

# Designing Monitoring and Evaluation Approaches under the Open Standards

## An FOS How-To Guide



November 2017 – **DRAFT - UNDER REVIEW**

**Foundations of Success**

*Improving the Practice of Conservation*

[www.FOSonline.org](http://www.FOSonline.org)  
[info@FOSonline.org](mailto:info@FOSonline.org)

This guide provides high-level guidance on designing monitoring and evaluation approaches for conservation projects and programs. It is based on the Adaptive Management principles and practices in the Conservation Measure Partnership's *Open Standards for the Practice of Conservation*.

To provide feedback or comments, you can contact Foundations of Success at [info@FOOnline.org](mailto:info@FOOnline.org). Visit our website at [www.FOOnline.org](http://www.FOOnline.org) to download the most recent version of this guide. Please cite this work as:

Foundations of Success. 2017. *Designing Monitoring and Evaluation Approaches under the Open Standards: How-To Guide*. Foundations of Success, Bethesda, Maryland, USA.



This work is licensed under the Creative Commons Attribution-NonCommercial-Share Alike 4.0 License. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-sa/4.0/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, CA 94105, USA.

Under this license, you are free to share this manual and adapt it for your use under the following conditions:

- You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).
- You may not use this work for commercial purposes.
- If you alter, transform, or build upon this work, you must remove the FOS logo, and you may distribute the resulting work only under the same or similar license to this one.

## **Contents**

Purpose of FOS How-To Guides .....	1
Overview .....	1
Defining Terminology .....	3
Preview of Key Decisions.....	3
Defining Your Broad M&E Needs .....	5
1) Identify your key audiences.....	6
2) Define whether you want to monitor and evaluate process or impact.....	7
3) Determine whether you will use your M&E for formative or summative purposes .....	10
4) Determine who will undertake the monitoring / evaluation .....	11
What Type of Evaluation is Most Appropriate for Adaptive Management?.....	11
Defining Your Comparison Model .....	12
5) Determine if you will try to compare those affected by your intervention with those not affected.....	12
Determining What and How You Will Monitor .....	15
6) Identify what (indicators) you will monitor .....	15
7) Consider what type of data you will collect .....	16
8) Determine how will you choose your subjects or entities to be monitored? .....	17
9) Decide the timing of observations and how many you will make.....	19
Concluding Remarks.....	22
Some References.....	23



## Purpose of FOS How-To Guides

This guide is one in a series of how-to guides designed to help conservation practitioners using the *Open Standards for the Practice of Conservation* operationalize their strategic plans. These guides are stand-alone documents, but practitioners will get the most value out of them when they use them together to support the broader process of moving from planning (Step 2) to implementation (Step 3).<sup>1</sup> The series of guides includes the following (relationship to Open Standards steps included in parentheses):

- A. Designing Monitoring and Evaluation Approaches (Step 2B, this guide)
- B. Data Collection Methods (Step 2B, under development)
- C. Operational Planning (Step 2C, under development)
- D. [Developing High-Level Work Plans and Budgets](#) (Step 3A)
- E. Funding Sources and Proposals (Step 3B, under development)
- F. Implementation (Step 3C, under development)

FOS staff members are developing these guides over time. As of May 2017, FOS had released drafts guides for topics A and D listed above. As the guides are published, they will be available on the [FOS website](#) and the [Open Standards website](#) (along with a peer-reviewed rating). The Open Standards website also contains implementation and operationalization guidance from other organizations, with Bush Heritage Australia providing numerous documents and examples based on their own experiences.

## Overview

Now that you have your indicators, you may feel that you are ready to go out and start collecting data. You are close to that point, but before doing so, it is helpful to take a step back and think about your monitoring and evaluation (M&E) design, as your M&E design will affect the degree to which you can attribute any observed changes to your conservation efforts.


In this guide, we share many concepts and terminology related to M&E design. We define M&E design as the approach a team takes to structure monitoring and/or evaluation – including sampling methods, use (or not) of controls and comparisons, timing of interventions, and timing of observations (Margoluis, Stem, Salafsky, & Brown, 2009). You and your team may find it useful to review these concepts and take an iterative approach to developing your M&E design, being careful not to become overwhelmed by your monitoring efforts or to lose sight of your overall project.

---


<sup>1</sup> For guidance on Steps 1 and 2 of the Open Standards, see FOS’s manual, *Conceptualizing and Planning Conservation Projects and Programs*, as well as FOS How-To Guides on conceptual models and results chains (available from [www.fosonline.org](http://www.fosonline.org)).

M&E design is important because it helps your team set up how it will do monitoring. For example, if you need a high level of certainty that your efforts are causing desired results, then you should consider comparing what is happening in the area where you are working with what is happening in a similar area where you are not working. You will also likely want to measure key variables before, during, and after you take action. If you do not need this high level of certainty, perhaps you would choose to look only in the area where you are taking action and then look at key variables before and after you take action.

You may not need to monitor every strategy in your plan; for some, you may even decide to schedule a periodic discussion and informal assessment among team members. All of these approaches are completely acceptable, but which approach your team chooses will be informed by various decisions that help inform your M&E design. This guide walks you through those various decisions

 **Tip!** The main purpose of M&E is to understand what is working and what is not - and to learn and adapt. Try not to get caught up in terminology - focus on the decisions your team needs to make to be able to determine what is working and what is not.

This guide intends to help you and your team understand some terms and concepts you may have heard. However, in presenting this information, we are not suggesting that all teams have to set up the most rigorous designs that will help them know with certainty that their strategies are leading to the desired changes along their results chain (theory of change). Indeed, in many conservation projects, this is rarely possible or desirable. The most important thing for you to keep in mind is that the main purpose of monitoring and evaluation is to understand what is working and what is not so that you can adapt and improve, as needed.

 **Tip!** Most conservation projects have limited resources, so most need to choose M&E designs that match those resources and that help them make good management decisions.

The topic of M&E design could take up an entire course. However, this section provides a basic overview of the concepts with which you need to be familiar in order to carry out or oversee successful monitoring and evaluation. It is not meant to be a complete description of everything you need to know and do, but it should help you better apply the process of adaptive management. The biggest decisions your team

will have to make are: 1) thinking about your audiences and their information needs; and 2) determining the level of precision needed to make management decisions.

## Defining Terminology

Keeping in mind that this manual is designed to help teams use the Open Standards to practice adaptive management, our focus is on monitoring to evaluate, learn from, and improve projects over time. We define “monitoring” as the periodic collection and analysis of data related to stated goals and objectives (Box 1). *Monitoring* allows teams to generate the data that facilitates *evaluation*. This distinction is consistent with the way the terms are used in the evaluation field. We primarily use the term “monitoring and evaluation (M&E)” to broadly encompass any effort to measure progress and impact.

In the conservation community, however, different actors use these terms differently, and you may run into inconsistencies. For example, some people consider “evaluation” as a formal process (typically carried out by a third party) to assess the progress of a project, program, or strategy. Some people also use terms such as “effectiveness” and “impact” to distinguish what types of data they are collecting. We will discuss this in more detail in a later section. The main thing to keep in mind is that, if you want to understand how effective your conservation efforts are and if you want to learn and improve over time, then your team should be measuring results along your stated theory of change – from beginning to end – regardless of whether you call that “effectiveness monitoring,” “impact evaluation,” or some other term.

Along the same lines, the specific indicators your team uses for testing progress along a results chain will usually be the same regardless of how you intend to use the data generated (e.g., whether you use the indicators for measuring short-term impact or long-term impact or whether you use them for project team learning or external assessment). For this reason, it is important that you develop indicators closely tied to your results chain (theory of change) during the strategic planning phase of your project.

## Preview of Key Decisions

The biggest, broadest decisions your team will have to make are defining your audiences and their information needs and the purpose of your M&E. For example, you may be conducting M&E to report project results to donors, your own institution, or any other stakeholder. These

### Box 1. Defining Key Monitoring and Evaluation Terms

**Monitoring:** The periodic collection and analysis of data relative to stated project goals and objectives. *Monitoring* enables you to generate the data necessary to *evaluate* the impact of your project. Some also refer to this process as *Monitoring and Evaluation (M&E)*.

**Monitoring and Evaluation Design:** The approach taken to structure monitoring and/or evaluation. It includes sampling methods, use (or not) of controls and comparisons, timing of interventions, and timing of observations (Margoluis, Stem, Salafsky, & Brown, 2009)

**Sampling:** The selection of a subset of individuals or entities from a population in order to estimate characteristics of the whole population.

Box 2 provides further definitions related to categories of M&E (process and impact).

audiences will want to see evidence that you are achieving what you said you would. But, if you wish to improve your project or future efforts, you should design your M&E to understand what works and what does not work (and, ideally, why). You could also be more formal with your M&E and seek to test assumptions and share your learning beyond your team. More likely, your team will orient M&E efforts to achieve a combination of all the above – but, this combination could be different for each strategy. For example, imagine your team is implementing a strategy with a key assumption upon which not all team members or stakeholders agree. In this case, you might want to take a rigorous approach to test that assumption and answer a key question (e.g., Under which conditions do economic benefits lead to desired environmental practices?). Your M&E results would provide important learning opportunities for your team but also for a broader audience. For another strategy, you might want to focus your M&E on helping you improve daily management decisions. For example, if you are implementing a law enforcement strategy, you probably do not need to determine if law enforcement decreases illegal activities. But you may want to focus your M&E on assessing which law enforcement activities are more effective and cost-efficient.

Thus, clearly defining the purpose of the M&E for each strategy (i.e., Why are you doing M&E? For whom are you doing it? and How are you expecting to use it?) will determine the level of precision you need. These decisions cut across several more specific decisions that we summarize here. The remainder of this guide provides more detail and guidance on each of these decisions and the implications for M&E design.

**Defining Your Broad M&E Needs:** When doing any monitoring, you need to:

- 1) Identify your key audiences and what they want to know
- 2) Define whether you want to evaluate
  - a. Process
  - b. Impact
- 3) Determine whether you will use your M&E information for
  - a. Formative purposes
  - b. Summative purposes
- 4) Determine who will undertake the monitoring / evaluation (which may not always be the same as who is collecting data on specific indicators):
  - a. Internal / first party
  - b. External / third party

These first four questions need to be answered before you can start thinking about your M&E design and what types of data you collect.

**Defining Your “Comparison Model”**



- 5) Determine if you will try to compare individuals, entities, or elements affected by your intervention with those not affected
  - a. No comparisons – non-experimental
  - b. Comparison groups – quasi-experimental
  - c. Control groups – experimental

Your team will probably use different comparison models for different indicators and/or questions.

**Determining What and How You Will Monitor:** Once you have answered the previous questions, you can then:

- 6) Identify what you will monitor – These are your indicators – a topic that is covered in-depth in Section 2B (Week 10) of [FOS’s Manual on Open Standards Steps 1 and 2](#).
- 7) Consider what type of data will you collect
  - a. Quantitative
  - b. Qualitative
- 8) Determine how you will choose your subjects (entities to be monitored)
  - a. Census
  - b. Sample
- 9) Decide the timing of observations and how many you will you make
  - a. Before implementation
  - b. During implementation, and/or
  - c. After implementation

In the following sections, we provide more detail on each of these topics to help your team make M&E design decisions that best meet your needs. These decisions will inform your choice of data collection methods. At the same time, you will probably need to revisit many of these topics as you are choosing your data collection methods (see Step 2B in FOS’s manual, [Conceptualizing and Planning Conservation Projects and Programs](#)) and make sure you are clear about the design you are using for each indicator, strategy, and/or project.

## **Defining Your Broad M&E Needs**


The first four decisions you need to consider when structuring your monitoring and evaluation work are:

- 1) Identify your key audiences
- 2) Define whether you want to evaluate process or impact
- 3) Determine whether you will use your monitoring information for formative or summative purposes
- 4) Determine who will undertake the overall monitoring

You will likely consider these four decisions simultaneously, and your decisions for each will probably influence the others.

**1) Identify your key audiences**

This is a critical step that you should have already done when you developed a first iteration of your monitoring plan under Step 2B of the Open Standards (see also FOS’s manual, [Conceptualizing and Planning Conservation Projects and Programs](#), Step 2B. Develop a Formal Monitoring Plan).

 **Tip!** Don't forget to include your own team as a key audience. Good adaptive management involves systematic M&E to test assumptions, adapt, and learn - so your team is a primary audience.

At this point, it is helpful to revisit your identified audiences and be clear about what information they want to know and what they need in order to do their job better. Giving thought to

**Box 2. Conditions Warranting High Investment in M&E Design**

A rigorous M&E design may make sense if:

- Project investment is high;
- An important audience needs strong inference of causality;
- Little is known about a strategy’s potential effectiveness;
- Project is a pilot that might be scaled up; and/or
- The team plans to publish project results in a peer-reviewed publication

the final product you will share with these audiences will help you determine what monitoring design and methods are most appropriate for your situation (Box 2 provides some examples of when a team may want to invest more heavily in M&E design). For example, if it is really important for your team to publish data on a specific project approach in a peer-reviewed academic journal, you will want to have as strong of a design as feasible, and you will want to use well-known and validated methods. Or if you are in charge of a large, well-funded project where you need to demonstrate with a high level of certainty the degree to which your project is having an impact, you should use a more rigorous design. If, however, your project is tight on resources and your

main audience is your project team and managers, you may choose a less rigorous design that will give you solid data that are sufficient for management decisions. For now, you can fill out the rest of your audiences table, which will help you make decisions related to selecting your monitoring design and methods.

**Table 1. Complete Table of Key Audiences and Information Needs for Marine Reserve Site**

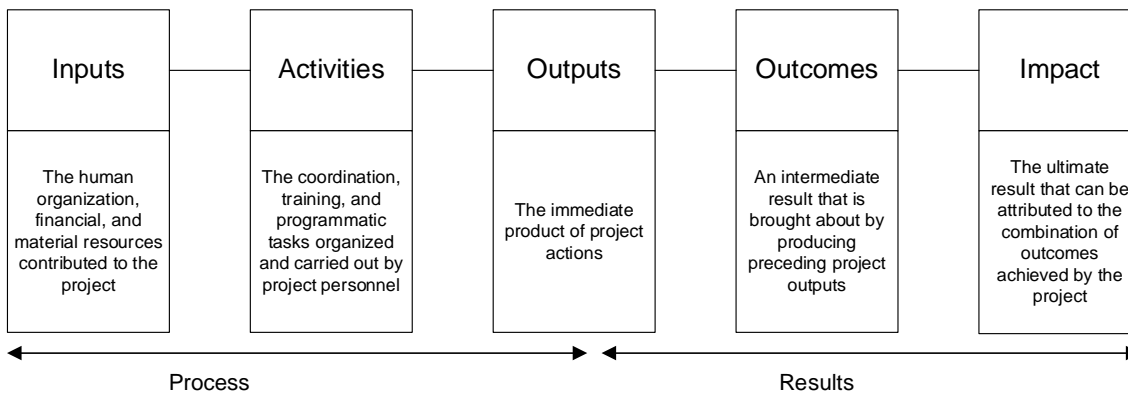
Audience	Indicators of Interest	Media Type/ Length	Desired Action
Project team	All indicators along results chain & those related to critical assumptions/ enabling conditions	Matrix of indicators & measurements, accompanying analysis by indicator	Learn, improve, adapt
Project partners	All indicators along results chain & those related to critical assumptions/ enabling conditions	Report with main results and conclusions by indicator, 20 pages (text heavy OK)	Learn, improve, adapt Support project and any proposed changes based on data

Audience	Indicators of Interest	Media Type/ Length	Desired Action
Save the Hammerheads	Presence/ absence of hammerhead sharks in key sites		
Senior managers	Key indicators along results chain (e.g., threat reduction, target status, & a few select indicators along the chain)	Dashboard of key indicators and associated analyses Anecdotes and/or hard data illustrating key achievements and challenges	Support project and any proposed changes based on data
Funders	Key indicators along results chain (e.g., threat reduction, target status, & a few select indicators along the chain)	Dashboard of key indicators and associated analyses Anecdotes and/or hard data illustrating key achievements and challenges	Support project and any proposed changes based on data Provide ongoing funding in the future
General public	Key indicators along results chain (e.g., threat reduction, target status, & possibly a few select indicators along the chain)	2 page glossy brochure 5-10 minute professionally produced video with key project highlights – including lessons learned	Support project and any proposed changes based on data

**2) Define whether you want to monitor and evaluate process or impact**

As shown in Figure 1, there are two main elements that a team may choose to monitor or evaluate: process and impact (or, more generically, results).

**Figure 1. Main Components of an Evaluation**



*Note: Some process evaluations might include outputs, as these are closely tied to activities. Some impact evaluations may also include some limited outputs, especially if their achievement represents important progress.*

**Process M&E** answers the question: Is the project doing what it said it would do – and is it doing these things well? Process M&E examines the implementation of project activities, procedures, and/or administrative and management functions. For example, a process evaluation could examine whether the marine reserve project team followed its work plan, met key deadlines for products, and did a good job of involving stakeholders in its activities.

**Impact (or Effectiveness)<sup>2</sup> M&E** answers the question: To what extent is the project achieving what it set out to achieve? Impact M&E examines whether results have been achieved over the short-term (outputs, to a limited degree, and intermediate outcomes) and the longer term (impacts). In the context of the work your team has done, this means measuring impact at key points along your results chain (or theory of change). An impact evaluation of the marine reserve project might examine whether the team was effective at achieving key results like getting local tourism operators to hook up to municipal wastewater systems, influencing policy makers to strengthen fisheries regulations, and responding to an oil spill near a fragile beach. As part of impact M&E, the team may want to look not just at whether the results were achieved, but is there a plausible case for one result causing another result (i.e., looking at the relationship between two or more results). A real-world example of impact M&E can be found in Rare’s approach to its Pride Campaigns, where they use the same theory of change across their social marketing efforts to collect data and test whether they are seeing the social and biological shifts needed for sustained conservation results (<https://www.rare.org/pride>).

Not all people use the term “impact” the same. Indeed, there are many definitions and interpretations of several terms. While it is most important for your team to focus on measuring change along your results chain, you should also be aware of these terms and how they are sometimes used. Box 3 shares some common (and sometimes conflicting) interpretations.

---

<sup>2</sup> See Box 3 for varying definitions and interpretations of the term “impact M&E.”

### **Box 3. Common Interpretations of Terms to Describe Impact M&E**

Different people, organizations, and fields use terms differently. We strongly suggest you and your team focus on the concept of monitoring key results along your results chains and not worry about terminology. However, it is good for you to know what these terms are and the different interpretations that exist.

#### **Impact Monitoring / Evaluation:**

*Interpretation 1:* Examines whether desired results along a results chain are being achieved – from intermediate results to threat reduction results to target results (interpretation used in this manual; synonymous with Interpretation 1 of performance M&E and effectiveness M&E)

*Interpretation 2:* Focuses on the ultimate impact (usually on a conservation target but sometimes on a threat reduction result; synonymous with Interpretation 1 of status M&E)

*Interpretation 3:* An evaluation with a rigorous design (experimental or quasi-experimental) with the aim to establish causality

#### **Performance Monitoring /Evaluation:**

*Interpretation 1:* Examines whether desired results along a results chain are being achieved – from intermediate results to threat reduction results to target results (synonymous with Interpretation 1 of impact M&E and effectiveness M&E)

*Interpretation 2:* Same as Interpretation 1 but focuses on intermediate results only

*Interpretation 3:* Same as Interpretation 1 but focuses on intermediate results only

*Interpretation 4:* Focuses on whether activities and tasks were carried out effectively and as planned (synonymous with “process evaluation,” as described in this manual).

#### **Effectiveness Monitoring / Evaluation:**

Used synonymously with Interpretations 1, 2, and 3 of performance monitoring / evaluation.

#### **Status Monitoring / Evaluation**

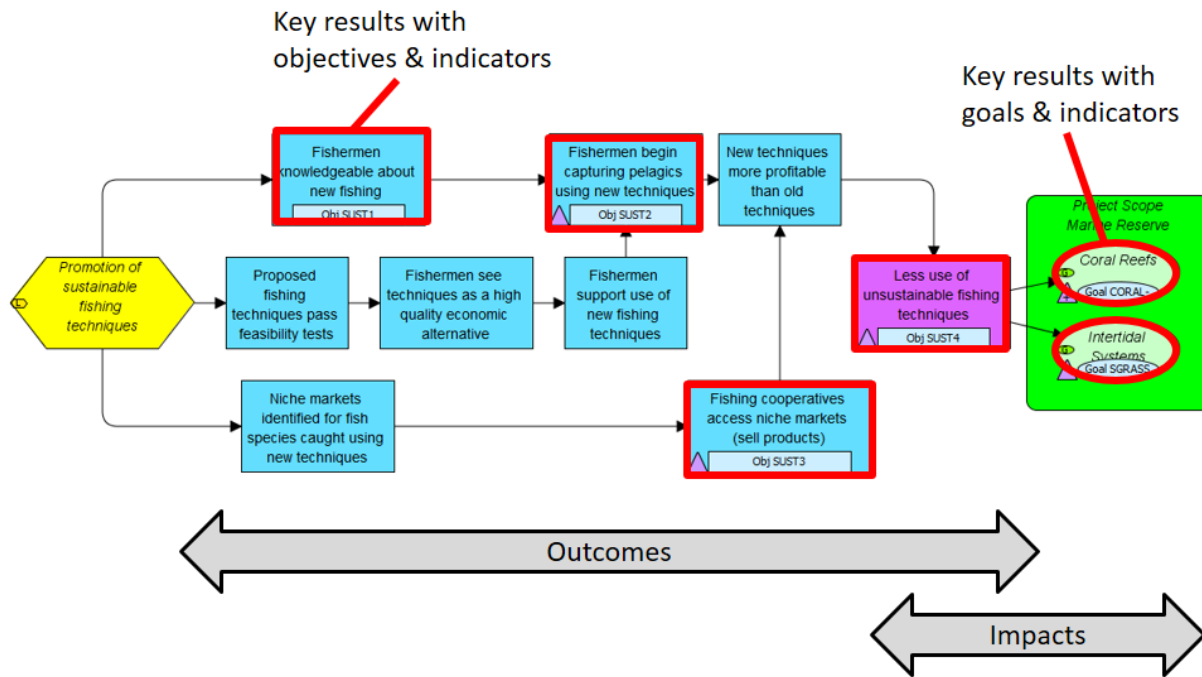
*Interpretation 1:* Focuses on the ultimate impact (usually on a conservation target but sometimes on a threat reduction result; synonymous with Interpretation 2 of impact M&E)

#### **Monitoring vs. Evaluation**

In addition to these distinctions, some people consider “monitoring” to be less rigorous and for internal purposes, while they see “evaluations” as more rigorous and used for accountability purposes. We see this distinction as somewhat artificial and not very informative – again, going back to our ongoing message that it is most important to use your results chain to measure impact along that chain and use that information to adapt and learn – regardless of what term you use to describe the process.

In monitoring for adaptive management purposes, the focus is typically on the results section of Figure 1. In a results chain and as shown in Figure 2, this means measuring impact at key results (and associated objectives), as well as longer-term impacts on conservation targets (expressed as goals under the Open Standards).

Figure 2. Main Components of Impact M&E



**3) Determine whether you will use your M&E for formative or summative purposes**

There are two main purposes that M&E may serve – formative or summative.

**Formative Evaluation** is done with the intent of gathering data solely or primarily to help improve projects and programs. Formative evaluation tends to happen on an ongoing basis.

**Summative Evaluation** is done to judge whether a project or program is performing as expected. Summative evaluations usually happen at key periods in a project’s life (e.g., midway through the project or at the end).


These purposes are not mutually exclusive. Indeed, the lines can be a bit blurry. The main distinction lies in the intent behind the evaluation and how the data are used. Because adaptive management involves learning and improving, M&E for adaptive management is done regularly throughout the life of a project for formative purposes. Nevertheless, good data can often serve both summative and formative purposes. If your team has done a good job laying out its theory of change and developing indicators tied to it, then the data you collect could serve both formative and summative purposes.

#### **4) Determine who will undertake the monitoring / evaluation**

Depending on data needs and resources, M&E might be conducted by an internal team, external team, or some combination of the two. The composition you choose will depend on how important it is for your results to be objective.

**Internal (or “First Party”) M&E** is conducted and managed by the same project staff that has designed and/or implemented the project.

**External (or “Third Party”) M&E** is performed by an outside evaluator who is independent of the project team and, therefore, seen as objective.

 **Tip!** Not all project staff have the skills to do M&E. You may have to contract outside support. However, all staff should be involved in setting the goals and objectives which will be monitored and orienting the external evaluator to your team's theory of change.

An evaluation could also include both internal and external representatives. In such cases, the internal members bring a rich understanding of the project, context, and history, while the external members help provide a more objective, removed perspective. Teams practicing adaptive management will usually do first party (internal) M&E because the main purpose of the monitoring is for the team to learn what is working and

what is not working. There is often not a need for an external, third-party perspective or the resources to bring in that perspective. However, some teams may have mixed composition – especially in cases where individuals on a project team do not have all the skills they need to carry out M&E.

#### ***What Type of Evaluation is Most Appropriate for Adaptive Management?***

As you can imagine, there are a number of ways that a team could structure its monitoring and evaluation work. For example, they could choose any one of the following combinations:

- Internal, formative, process evaluation;
- External, summative, impact evaluation;
- Internal, formative, impact evaluation
- External, summative, process evaluation
- Internal, summative, impact evaluation

This list is not exhaustive; any permutation is possible. However, it is most common that a summative evaluation would be conducted externally. In contrast, adaptive management tends to promote learning within a team to improve a project over time. As such, it is more typical that monitoring for adaptive management would focus on internal, formative, impact evaluation. Although strong evidence is important, practicing adaptive management does not require strict adherence to determining causality *at any cost*. Given the focus of the Open Standards on promoting adaptive management, this training manual focuses more on internal, formative

evaluation and does not emphasize the strict requirements often associated with external summative evaluations.

## Defining Your Comparison Model

Now that you have defined your broad M&E needs, it is time to start thinking about the approach you will use to test the assumptions in your results chain. We refer to this as your “comparison model,” though some people will call this an “M&E design.”<sup>3</sup> More specifically, this decision is about determining whether you will use comparison or control groups or whether you will examine only the population or entity you are trying to influence. This step may be influenced by later steps. While presented sequentially here, there will often be some back and forth between this step and later steps (see the [Determining What and How You Will Monitor](#) section).

### ***5) Determine if you will try to compare those affected by your intervention with those not affected***

There are three options your team should consider (listed in the order that is more common for on-the-ground conservation projects):

- a) No comparisons or controls – known in the research and evaluation world as a non-experimental design;
- b) Comparison groups – known as quasi-experimental design; and
- c) Control groups – known as experimental design.

Your team will probably define a general comparison approach for your M&E efforts.

However, you may also decide to use a different comparison model for a particular strategy or even individual results and indicators. For example, you may decide to not compare your site or project to another site or project, but you may decide to look at your project’s ability to affect a specific indicator (e.g., forest cover) as compared to values of that indicator at multiple non-intervention sites because this is data you can easily obtain across various sites. In general, most small or medium-sized conservation projects will not have the resources to use comparisons or controls. Box 4 provides some general

#### **Box 4. General Conditions for Determining Your Comparison Model**

Your monitoring audience and information needs ultimately should inform which comparison model is most appropriate for your situation. However, some general principles include:

Consider a **less rigorous approach** when:

- You are using a strategy with a good evidence base showing its effectiveness
- Investments are relatively small
- Risks to conservation and/or society are low if strategy is not effective

Consider a **more rigorous approach** when:

- This is a new strategy and/or the evidence base demonstrating the effectiveness of this strategy is weak
- Project or strategy is a pilot that might be scaled up
- Investments are high
- Risks to conservation and/or society are low if strategy is not effective
- You have an obligation to one of your

<sup>3</sup> As mentioned earlier, we define “M&E design” as broader than the comparison model. It includes: sampling methods, use (or not) of controls and comparisons, timing of interventions, and timing of observations (Margoluis, Stem, Salafsky, & Brown, 2009). For this reason, we do not use the general term of “M&E design” to characterize the use (or not) of controls or comparison groups.



considerations that may influence the type of comparison model that is most appropriate for your team.

### Box 5. Qualitative Approaches

Many people use the terms “qualitative methods” and “qualitative design” loosely (and even interchangeably). While some people categorize qualitative design as another type of design, it is really a type of non-experimental design. Qualitative approaches may involve one or more of the following elements:

**Qualitative Sampling:** Focuses on the sampling framework and not statistical power. Individual cases are weighted more heavily because the evaluator is not looking for population-based trends. Some qualitative sampling examples include stratified purposeful sampling and theory-based sampling.


**Qualitative Methods:** Focus on extracting a limited amount of rich data that provide contextual information to understand observed patterns and relationships. Examples include key informant interviews, focus groups, direct observation, and participatory mapping.

**Qualitative Data:** Describe qualities of whatever is being studied but are not numerical in nature. Examples include informant transcriptions or recordings, videos, oral histories, and field notes. Within these, the evaluator might be trying to be identifying more specific data, such as perceptions about wildlife abundance (e.g., there many fewer elephants), opinions about the environmental regulations (e.g., regulations are strict), and descriptions about water quality (e.g., the water is murky).

**Qualitative Analyses:** Examining (usually qualitative) data to understand and explain the reason behind behaviors or conditions. While there is no generally no formula for doing qualitative analyses, the evaluator is often looking for patterns and commonalities or differences that can help explain those patterns. Examples include discourse analysis, coding data (based on themes), and iterative or grounded analysis.

In general, qualitative approaches try to understand the context from the perspective of the informant(s). They provide a richness that complements quantitative approaches, often aiding understanding the “why” behind observed trends.


a) **No comparisons or controls (non-experimental design):** This comparison model does not involve control or comparison groups and, as such, it is the least expensive approach. It is also

 **Tip!** Any of the approaches laid out here can be strengthened by taking multiple observations over time (see [Step 9](#), described later in this document).

the most common approach used for on-the-ground conservation projects. This approach requires less time and resources to implement, but it also has lesser ability to detect causal relationships. Non-experimental designs may rely on quantitative approaches and/or qualitative approaches (see Box 5 for a description of qualitative approaches).

b) **Comparison groups (quasi-experimental design):** This approach involves identifying comparison groups which will not participate in or benefit from a project or strategy. The project team tries to find comparison groups that are as similar as possible to the group “exposed” to the project or strategy. This allows the team to develop a counterfactual situation – what would have happened in the exposed group if they had not been exposed to the project or strategy. It is


similar to experimental designs (described below) but lacks the random assignment of subjects to the comparison group. A common example of a quasi-experimental design is matching, where an exposed group is matched with a non-exposed group selected by the team. For instance, suppose you are working to promote sustainable fishing practices, and you want to understand whether fishers’ use of these practices is greater in communities in which you have worked. You should measure fishers’ use of the practices in those communities. You could then also identify other similar communities in which fishers have not participated in your strategy, and you could measure those non-exposed fishers’ use of the practices.

 **Tip!** Any evaluation design (including experimental design) will be strongest when it includes a mixture of quantitative and qualitative methods.

Another example of a quasi-experimental design which can be quite powerful is a cross-site comparison. Imagine now you are working in three different sites where you are promoting sustainable fishing practices – all fishers have been exposed to the strategy. However, some variables that differ across those sites include geographic region, community organization, and income levels. Your team could try to assess whether there are differences in fishers’ adoption of sustainable fishing practices across the three sites and whether geography, community organization, and/or income levels might influence adoption. In this way, you are not formally testing the effectiveness of your efforts to promote sustainable fishing practices (because you are controlling for this variable by ensuring all are exposed). Rather, you are testing the degree to which other key variables (differences among communities) may affect adoption rates.

The main advantage of using comparison groups is that doing so can be more feasible to set up than a true experimental design, yet it yields strong evidence for causality. Quasi-experimental designs, however, tend to be moderately to highly expensive and, as such, are often not practical for conservation teams.

c) **Controls (experimental design):** This approach involves the random assignment of subjects to those “exposed” to the strategy or project (treatment group) and those “not exposed” (control group). The main advantage of an experimental design is that it provides strong evidence for causality. However, it is also very expensive and typically not practical for most conservation teams.

 **Tip!** Teams can often feel pressure to provide strong evidence of causality and, therefore, to use more rigorous approaches. While it is ideal to provide reasonable evidence of causality, it is important to remember that conservation happens in a dynamic, messy, resource-limited context. See Box 4 for more tips.

You will need to think about which design is right for your needs. Some key considerations will be the resources available to you for monitoring and the level of certainty you need in order to feel confident that your project is on the right track. If you have a lot of resources for monitoring and you need a high level of certainty about causal relationships, you might want to choose an

experimental or (more likely) a quasi-experimental design. More typically, however, resources available to conservation teams are more limited and, and as such, non-experimental designs are more commonly used. If your resources are more restricted, but you still have a need to show fairly strong causality, you might choose a non-experimental design. If you do so, you should ideally be collecting data on your indicators before the project begins and after you expect to see changes as a result of your conservation actions (i.e., pre-test/post-test or time series design). Finally, qualitative designs might be appropriate if you do not need to show causality, but you want to highlight rich stories and experiences. Qualitative designs may be less costly to implement, but you should be sure to consider data analysis and interpretation costs when considering the overall cost of a qualitative design.

## Determining What and How You Will Monitor


Once you know your broad M&E needs and your general “comparison model,” you can then turn to the substance of the monitoring itself. In particular, you will need to think about the following decisions:

- 6) Identify what you will monitor
- 7) Consider what type of data you will collect (quantitative, qualitative, or both)
- 8) Determine how you will choose your subjects (entities to be monitored)
- 9) Decide the timing of observations and how many you will you make

These decisions, as well as the decision about your comparison model, will be heavily influenced by your M&E aims and resources. Remember that these steps can be iterative, so you may move through Steps 6-9 and then go back to determine or refine your comparison model.

### **6) Identify what (indicators) you will monitor**

If you have been following Open Standards guidance on results chains, objectives, and indicators

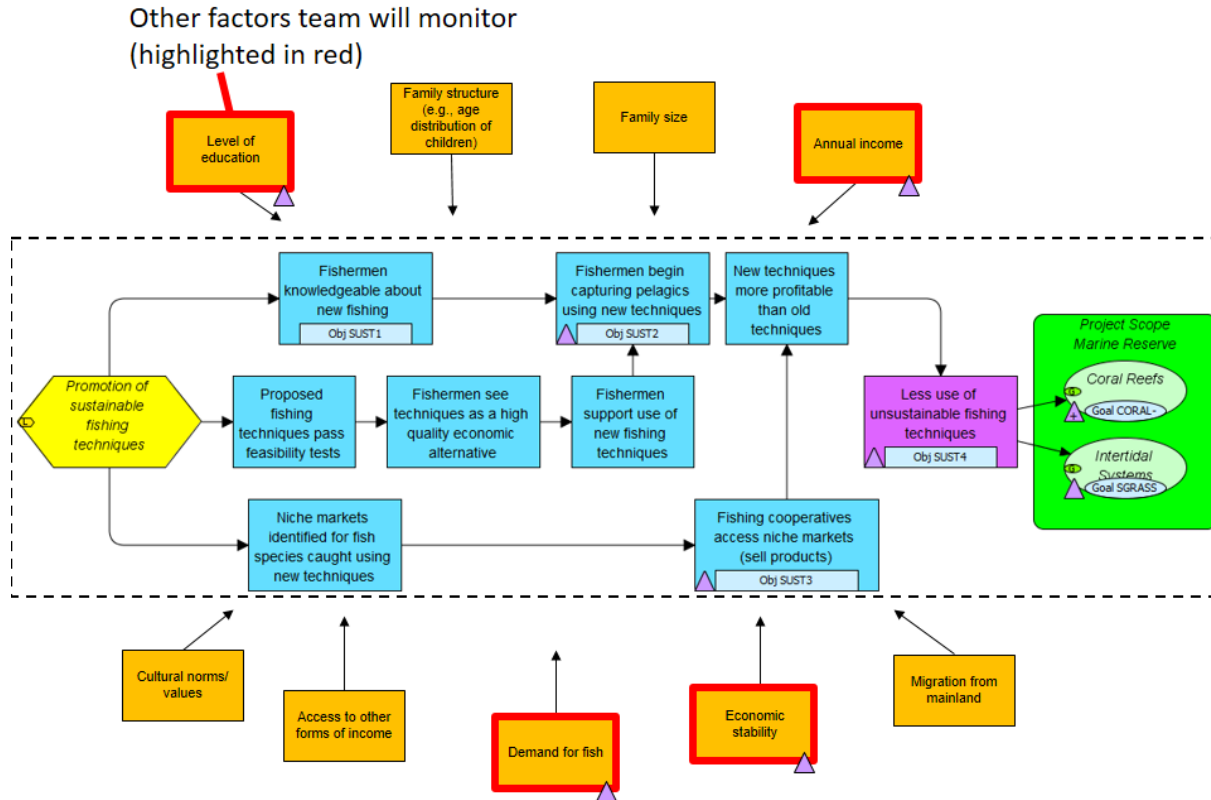
 **Tip!** If you monitor other factors outside of your results chain, try to limit those factors to ones with a strong likelihood to influence (positively or negatively) your ability to achieve your results. While it is important to try to establish causality, you also need to keep your overall monitoring to a reasonable level for your budget.

(see [FOS Manual, Step 2A, Weeks 9-11](#)), your team should have identified and be prepared to measure key results (and associated objectives and goals) along your results chain (refer back to Figure 2 for an example). In addition to these, you also may want to monitor any other factors or issues not expressed in your results chain but that may affect your ability to achieve your goals and objectives. Doing so will help increase your ability to demonstrate potential causality. A good starting place is to look at the conceptual model you developed in Step 1 of the Open Standards. You may want to consider contributing factors that you did not bring over to your results chain but that were causally linked in your

conceptual model (e.g., migration from mainland). In addition, there may be some broad issues (e.g., political stability, economic trends) or characteristics of target audience (e.g., education and income) that you may want to monitor because they could influence your ability to achieve

your results (Figure 3). Your challenge will be finding the right balance between what you need to monitor and what might be nice to monitor.

Figure 3. Other Factors to Consider When Evaluating Project Impact



### 7) Consider what type of data you will collect

There are two main types of data: quantitative and qualitative.

**Quantitative data** are data that can be measured in numbers or percentages and which can be put into categories or in rank order. This type of data is often presented in graph or tabular form.

**Qualitative data** are in-depth descriptive data that are typically observed but not measured in numbers or percentages. Examples include feelings, opinions, stories, and observations.

The following table summarizes the main advantages and disadvantages of each type of data. In general, the strongest monitoring efforts include a mix of quantitative and qualitative data (and methods).

**Table 2. Quantitative and Qualitative Data – Main Advantages and Disadvantages**

	Advantages	Disadvantages
<b>Quantitative Data</b>	Easy to analyze Allow for broad generalizations across the sample	Does not provide contextual background helpful for analyzing responses
<b>Qualitative Data</b>	Provide in-depth description, detail, and/or richness High validity	Difficult to analyze Difficult to generalize

However, keep in mind that here we are discussing quantitative and qualitative *data*. Some people equate quantitative data with quantitative methods and qualitative data with qualitative methods. There is a high degree of overlap, but it is not complete (Box 5). For instance, quantitative methods such as questionnaires (typically considered a quantitative method) can produce both quantitative and qualitative data. More specifically, a survey might collect close-ended responses which could be put in categories (e.g., “yes” or “no” answers, or 1 to 5 ratings). However, it might also include open-ended questions that would generate qualitative descriptive data. Likewise, key informant interviews (a qualitative method) could include questions seeking quantitative data (e.g., budget figures or number of projects exhibiting certain characteristics). See Box 5 and **X guide**, in particular, for more detail on quantitative and qualitative *methods*.


**8) Determine how will you choose your subjects or entities to be monitored?**

You also need to think about on which individuals, units, or entities within a population you will collect data. There are two options:

**Census monitoring** involves measuring all the individuals, units, or entities in a population. Using a census has the inherent advantage that you know your monitoring data will adequately represent the entire population since you have collected data for all individuals! In some projects, you can easily monitor all the individuals in a population. For instance, if you are interested in monitoring change in environmental knowledge and attitudes among students in an urban classroom participating in an environmental education course, it would be fairly easy to administer a survey to all the students. Likewise, if you are interested in measuring the number of fruit trees in a small agroforestry plot, you can count all the trees in the plot. Or, if you want to know the number of governments that have signed on to a treaty, you could easily collect that data. In addition, there may be automated ways of gathering data from an entire population. For example, you may be interested in how many visitors to a website download a certain publication. Even if the population of visitors is quite large, you could use Google Analytics or other tracking software to generate the desired data with a few mouse clicks.

**Sampling** involves measuring a representative subset of individuals, units, or entities in the population. Suppose in the previous examples that the populations you want to monitor are much larger – for example, you want to monitor students’ environmental knowledge and attitudes across 50 school districts. Or, perhaps you need to know how many trees of a certain species are in a 500,000 hectare forest. For either case, it is technically possible to survey all students or count all trees, but it would not be a very good use of project resources. By identifying a representative sample, you can monitor the desired indicator and then extrapolate the results to the broader population.

The main advantages of sampling over census monitoring lie in the reduced cost and greater speed of measurement because you are collecting data on a subset of individuals or entities rather than on an entire population. In our Marine Reserve example, it would be impractical to count every fish in the coral reef habitat. More likely, you will only be able to monitor selected samples. The disadvantage of sampling is it can be quite complicated to do it well, and it requires using statistics to determine to what extent your sample data represent the population. When selecting your sample, you will want to take care to minimize selection bias which can affect your ability to generalize conclusions.

 **Tip!** Use care in selecting your sample, as your ability to generalize your conclusions to the broader population is dependent upon the degree to which your sample is representative of that broader population. If you wish to make well-founded attribution or causality claims, you should consider working with an experienced statistician to identify the best sampling method.

There are different approaches to sampling, with different benefits and drawbacks. It is beyond the scope of this guide to discuss all the different approaches and how to best employ them.. Instead, we highlight, in very general terms, a few concepts. If you decide to sample, your team may wish to bring in the expertise of a statistician. One common approach to sampling your team may consider is a simple random sample – a sampling technique that minimizes bias and simplifies analysis of results. However, because of the random nature of selection, it does not allow you to easily examine a subset of the population that may be of interest. In this case, stratified sampling would be helpful because it allows you to stratify your sampling frame by key characteristics of interest (e.g., race or income level). Each strata then becomes your population from which you sample. Stratified sampling can be more costly and require a larger sample overall. In addition, it tends to work only where sub-groups are fairly homogenous.

In qualitative approaches, you might use sampling techniques such as purposeful sampling, where you seek out individuals or entities that will provide a lot of information about the topic of interest. For example, you might select individuals based on a set of criteria, such as female heads of household from fisher families. Maybe you wish to purposefully sample this group because they are the ones who will be able to give you the most accurate information about annual household income. Another example of a qualitative sampling approach is snowball

sampling, where you might interview key individuals and then ask them whom else you should interview, based on your topic of interest. For more information on quantitative and qualitative approaches to sampling, your team could consult academic textbooks and/or work closely with experienced professionals.

**9) *Decide the timing of observations and how many you will make***

The last key consideration is determining when and how often you will monitor. There are three options:

- a. Before implementation
- b. During implementation, and/or
- c. After implementation

In the evaluation literature, these are also referred to as: pre-test (before implementation), post-test (after implementation), and time-series (a number of observations over time that may span before, during, and/or after implementation). Determining the timing of measurements and how many you take is largely influenced by the level of certainty you need to know that any changes you witness are likely due to your conservation actions. The same considerations outlined in Box 4 apply in this step as well. Box 6 provides additional tips to consider for determining when to collect data.

### **Box 6. Considerations for When to Collect Data**

It is important to define the dates when baseline and final data will be collected for each indicator. In many cases, you will want to collect data more frequently. In deciding when and how often you should collect data, consider the following factors:

- **Time period to effect change:** If you realistically cannot expect to see a change in a factor for five years after the start of the project, then your next measurement after the baseline measurement should probably be no earlier than five years (unless you need to monitor it for the influence of other variables).
- **Seasonality:** You may need to always monitor at the same time of year or at various points of the year to be able to factor in seasonal changes. For example, if you are monitoring water levels, they will vary widely depending upon whether you take them at the beginning of the wet season versus during or at the end of the wet season. In most cases, it would not make sense to compare water levels taken at the end of the dry season one year with those taken at the end of the wet season the following year.
- **Natural variability:** If what you are monitoring varies naturally, you should have enough data points from appropriate timeframes to account for natural variations that have nothing to do with project impacts. For example, if you are collecting data influenced by climatic changes, you should clearly note if the measurement time coincides with an El Niño year and how that might affect your results. You may also want to vary the number of collection times around the El Niño event to compensate for this effect.
- **Project life cycle:** This is a more practical concern. You should keep in mind key project reviews, planning, reports, or other project-related events and adjust your monitoring times to meet those needs if doing so will not substantially affect the outcome of your monitoring.

The least rigorous option is to take a single measurement after implementing your conservation action (post-test). You can significantly increase your ability to demonstrate possible causality by taking measurements before and after implementation (pre-test / post-test). You can often improve that ability even further by taking several measurements over time because doing so could allow you to notice whether the variable of concern is sensitive to a particular time (e.g., seasonal changes or political cycles). Remember that more rigor may be unnecessary and an unwise use of resources. Your team will need to weigh its available resources against the need or desire to demonstrate causality before deciding the timing and number of measurements. Box 4 can help you think through this decision. Likewise, Figure 4 summarizes the main trade-offs your team should consider when determining the timing of observations and your comparison model.



Figure 4. Cost and Precision Trade-Offs of Comparing Treatment Groups and Timing of Observations

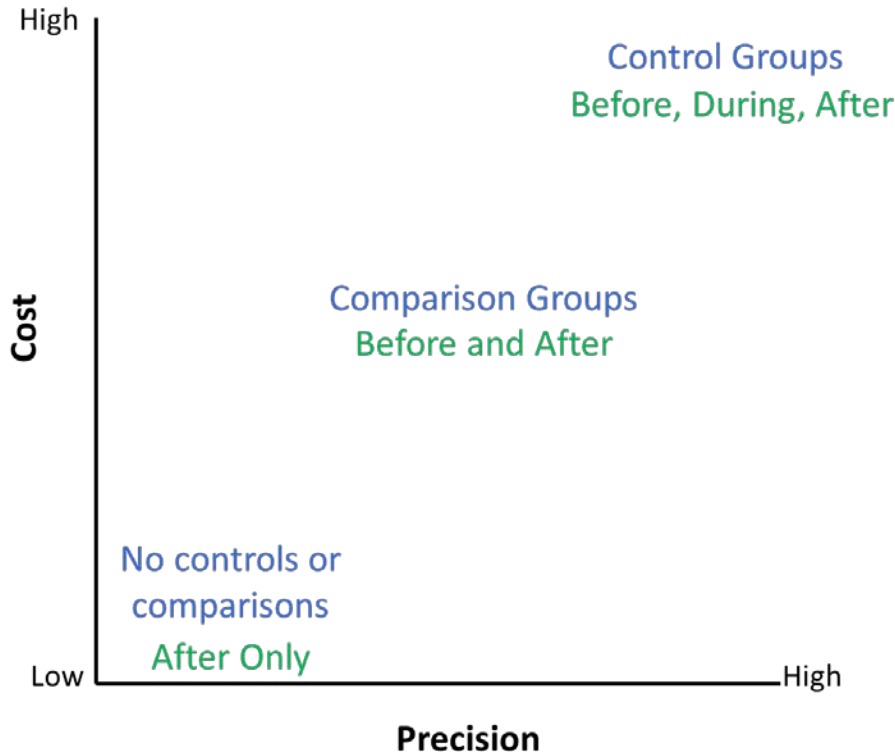


Table 3 summarizes some potential methods and comparison models that a team implementing a sustainable fishing strategy could use to collect data on its indicator. As you can see in the table, the comparison model might vary by indicator. In a real-world context, if the team were collecting indicator data for its own ongoing monitoring purposes, it may not use comparison groups, and it almost definitely would not use control groups. Likewise, the team would likely use post-test or possibly pre-test/ post-test observations. For data that are easier to collect, the team might do time-series observations

Table 3. Potential Methods, Samples, and Comparison Models for Indicators for Internal Project Monitoring for Sustainable Fishing Strategy

Indicator	Method	Sample/ Comparison Model
% of fishers in identified terrestrial ecosystems and watersheds that can name and describe at least two new sustainable fishing practices	Survey fishers	Sample of fishers Comparison with fishers in areas not participating in intervention Observations before, during, and after intervention
% of the fishers in identified terrestrial ecosystems and watersheds that are using only sustainable fishing practices	Review registry of fishing gear on boats before they leave on fishing expeditions	Sample of boats No comparisons or controls Observations before and after intervention

Indicator	Method	Sample/ Comparison Model
	Random periodic checks of fishing boats	
% of fishing cooperatives within identified terrestrial ecosystems and watersheds that are selling sustainably-caught fish to high-end markets	Interview managers at fishing cooperatives	Census all fishing coop managers Comparison with coops in areas not participating in intervention Observations before and after intervention
% of the fishers using the new practices that are earning at least 20% more income than under the unsustainable method	Survey fishers	Sample of fishers using the practices No comparisons or controls Observations before and after intervention
# of tons of key identified species caught	Review Fisheries agency’s records of fish landings Review Fishing Cooperative’s records of fish brought in for processing and sale	Census all records Comparison with fish caught outside intervention area Observations before and after intervention
% of the sub-watersheds classified as having “intact” river fish populations	Biological survey	Sample of sub-watersheds No comparisons or controls Observations before and after intervention

## Concluding Remarks

This guide is designed to help you operationalize your monitoring efforts. You have invested a lot of time and effort into developing your strategic plan, and you want to make sure that you track whether what you expected to see happens in reality. To do that, you need to make some key decisions about your audience, their information needs, and the purpose of your M&E. These decisions will influence your M&E design and the level of rigor and attribution you will need to consider to meet your key audiences’ information needs. You will also need to carefully choose your data collection methods. For more guidance on this topic, please see [FOS’s How-To Guide on Data Collection Methods](#).

To practice adaptive management, your team needs to have the right information available at the right level of detail to help you make good management decisions. This usually does not mean irrefutable proof your actions are working. Rather, it means having strong enough evidence to inform management decisions with a reasonable degree of certainty. What is “reasonable” will vary by projects, contexts, and actions. Investing some time upfront in M&E design will help you ensure you have the framework and data necessary to be able to analyze, learn, and adapt – go “full cycle” with the Open Standards.

## **Some References**

There are countless resources related to M&E design. The following are a few key, relevant resources that your team may find helpful. We encourage you to explore the evaluation literature (articles, textbooks, and guides) to find additional information tailored to your needs.

- Bamberger, Michael, Rugh, Jim, and Mabry, Linda. 2006. *RealWorld Evaluation: Working Under Budget, Time, Data and Political Constraints*. Thousand Oaks, London, New Delhi: SAGE Publications. [www.realworldevaluation.org/RealWorld\\_Evaluation\\_resour.html](http://www.realworldevaluation.org/RealWorld_Evaluation_resour.html).
- Margoluis, R., Stem, C., Salafsky, N., & Brown, M. (2009). Design alternatives for evaluating the impact of conservation projects. In M. Birnbaum & P. Mickwitz (Eds.), *Environmental program and policy evaluation: Addressing methodological challenges*. *New Directions for Evaluation*, 122, 85–96.
- Patton, Michael Quinn. *Qualitative Research & Evaluation Methods: Integrating Theory and Practice*. Thousand Oaks, CA: SAGE Publications, 2015.
- Rossi, Peter H., Mark W. Lipsey, and Howard E. Freeman. *Evaluation: A Systematic Approach*. Thousand Oaks, Calif.: Sage Publications, 2009.