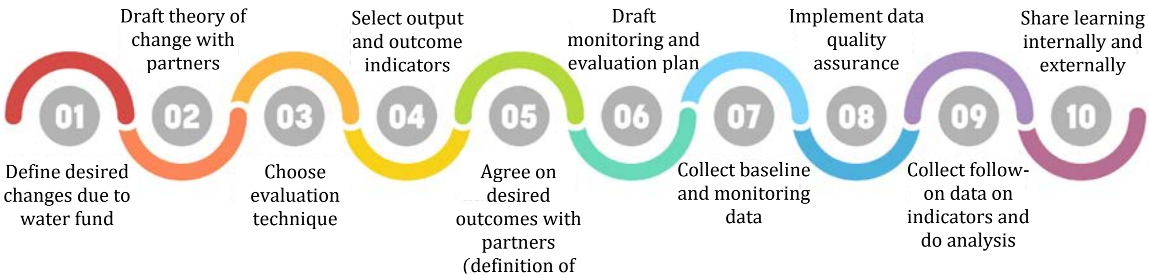
A Guide to Monitoring and Evaluating Water Funds

By Craig Leisher, David Courtemanch, Nathan Karres, Paulo Petry, and John Sowles



The Nature Conservancy
Protecting nature. Preserving life.

Contents

Acknowledgements

This guide is an output of The Nature Conservancy's water fund Monitoring Action Team led by Nathan Karres. The seed, however, came from Jonathan Higgins (along with a lot of laughter) who did the seminal work on water fund monitoring that is a key part of this guide. Colin Apse secured the funding for the Monitoring Action Team and guided the process. Kari Vigerstol was the overall manager, and Daniel Shemie challenged us with great questions. Anthony Kariuki, Claudio Klemz, Llara Kritzner Jackowski, Carolina Patricia Polania Silgado, Carlos Andres Rogéliz Prada, Gina Waterfield, and Johnny Wilson reviewed various drafts and provided helpful comments. Scott Chaplowe, an external monitoring and evaluation expert, provided a critical review and a number of helpful suggestions.

There are two different types of people in the world: those who want to know, and those who want to believe.

—Friedrich Nietzsche

Summary

A water fund's success depends on implementation, yet we rarely monitor water fund implementation or use adaptive management to improve our activities. Hence, our water fund successes are less than they could be.

Another missing element in most water funds is an evaluation process. Evaluations are critical for building the evidence to show that water funds make a difference for people and nature and for improving the design of new water funds.

The two most common reasons stated for not doing monitoring and evaluation are limited capacity and insufficient funding. Both are tractable problems. Capacity limitations can be addressed by building or buying the needed capacity as detailed in the main text. Insufficient funding can be addressed by including monitoring and evaluation in the budget of a new water fund or adding it to an existing water fund's budget. Neither need be expensive. Adding the minimum level of monitoring for a water fund is estimated to cost approximately US$18,000 a year plus upfront costs. To assist water fund teams in understanding what to do and how much it might cost, we provide a menu of monitoring and evaluation options in the main text with the options ordered by priority.

Every water fund would benefit from a monitoring and evaluation plan. Even a minimal plan is better than no plan (or dismissing monitoring and evaluation as nothing but a hope or a hinderance). When developing this plan, start with the desired changes and a theory of change that defines the hypotheses ("theories") about how desired changes can happens. The theory of change in turn informs the selection of output and outcome indicators. Stakeholders who are involved in the theory of change and indicator selection are more likely to own and support them. Because most water funds are partnerships, getting a consensus among partners on the outcome indicators is vital because the outcomes we choose to measure becomes how we define success.

Baseline data are critical for evaluating changes over time and providing a comparison point. Collecting baseline data is not as simple as it sounds, though. For water quality and quantity indicators where there is no historic data, baseline studies often need to start several years prior to on-the-ground activities to calibrate the relationship between water fund activity sites and comparison/control sites because baseline values may fluctuate seasonally or depend on water flow levels. Consider starting baseline data collection at the pilot stage of a water fund.

Socioeconomic baselines are challenging as well, especially identifying comparison/control and ensuring comparison/control benefit from participation. Evaluation experts in other sectors have found a stepped-wedge design approach helpful in addressing these issues. See the main text for more on this.

In sum, if we want to sustain existing water funds and encourage new ones, we need implementation monitoring and impact evaluations. Implementation monitoring maximizes the chance of a water fund being successful because it allows for adaptive management, and an impact evaluation generates evidence of benefits to people and nature and highlights what works.

This guide shows how to do both of these in a pragmatic way (and with a bit of humor).

Why this Guide

Water funds are a compelling idea, but after more than a decade of implementing them, we have little evidence that water funds truly benefit people and nature. To move from faith to fact, we need evidence, and this requires what is sometimes called "measures" but what is more widely (and accurately) known as "monitoring and evaluation."

This is a guide for monitoring and evaluation of water funds.[1] It is aimed at designers of new water funds and managers of existing water funds. It is a blend of monitoring and evaluation guidance, new ideas, and sage advice.

This guide is what one needs to know to do monitoring and evaluation for a water fund and avoid lethal monitoring and evaluation mistakes (text box). It is also short and thus does not include step-by-step instructions but adds hyperlinks to these instead. It's a compass rather than a cookbook.

Four Lethal Mistakes in Water Fund M&E

1. Poor communication between partners
2. Vague objectives and/or poor experimental design
3. Failure to act on and not just review the data collected
4. Failure to turn data into information

Why Monitoring Matters

Monitoring is the often overlooked and undervalued sibling of evaluations. Yet it matters more than evaluations for the success of water funds. Success depends on good implementation, and implementation monitoring is a proven way to improve project implementation.

- Conservation initiatives fail as much from poor implementation as from poor design. The only way to know if activities are implemented as planned, or could be implemented in a better way, is to monitor implementation—tracking inputs, activities, and outputs systematically. What gets measured, gets done.

- Water funds often use new approaches to address local issues and are implemented in complex social-ecological systems whereby when one social or ecological factor changes, other factors change as well. Adaptive management addresses the inherent uncertainties through learning. Adaptive management is about "managing to learn" and "learning to manage." Only with implementation monitoring can a water fund adaptively manage project implementation. Bad news can become good news if one does something about it.

- Replicating a project has little chance of success if the original inputs are unclear. Knowing the ingredients needed for replication is critical, and for this we need implementation monitoring.

In short, if water funds want to create large-scale, durable results, then implementation monitoring is needed. Without this, our successes will be diminished.

[1] This guide draws from the authors combined 124 years (!) of monitoring experience and pulls the best ideas from *The Goldilocks Challenge*, a review of monitoring and evaluation systems at 19 other international NGOs, and the document that started it all, the *Primer for Monitoring Water Funds.* Pithy points from the book *Applied Wisdom for Nonprofits* are woven in as well.

Sad Reality of Evaluations in Water Funds

Within conservation in general and water funds in particular, project evaluations are the exception rather than the norm. When they do happen, they are usually donor-driven and assess a project's implementation performance. Such evaluations are not without merit, but they say little about the impacts of the project. For this we need impact evaluations.

Impact evaluations measure what actually changed because of a project or activity. They are powerful tools for influencing funding flows, especially when they are quantitative. The number of international development projects with quantitative impact assessment has grown rapidly in recent years, thanks to organizations such as 3ie, J-PAL, and IPA. Quantitative impact evaluations have become a primary influencer of international donor funding.

Within quantitative impact evaluations, Randomized Control Trials are the dominant technique. For a variety of reasons, however, Randomized Control Trials are challenging for conservation projects. If Randomized Control Trials are not the answer, what is? The quantitative impact evaluation technique with the most promise for conservation projects is called Difference-in-Differences. This technique is increasingly used to measure conservation outcomes. The gist of Difference-in-Differences is that pre-existing differences between a project group or area (a.k.a. "treatment") and a control group or area are identified (the first difference) and these differences are subtracted from the results after treatment (the second difference). An overview of how a Difference-in-Difference technique works is here.[2]

Quantitative evaluations matter if we want to increase the success and use of water funds globally.

Defining the Desired Changes and Developing a Theory of Change

A water fund's monitoring and evaluation starts with a theory of change. A theory of change presents the hypotheses ("theories") about the desired changes a project or group aims to achieve.

There are a number of steps for creating a theory of change, and one of the first steps is defining the desired changes a water fund wants to achieve. Most water funds are partnerships, and it is important at this point to get consensus on the desired changes with the partner organizations. (And avoid lethal monitoring and evaluation mistake #1: poor communication between partners.) The more partner organizations are involved in the theory of change development, the more likely they are to support the results.

Another key step in creating a theory of change is to know your basin or watershed. Understanding the sources of the problems to be addressed is critical. One also need to understand when and where changes can be expected. For more on temporal and spatial scales and determining realistic goals for rehabilitating a basin/watershed, see the assessment framework here. For more on assessing stream condition and stability, see here.

[2] Difference-in-Differences is widely used but has to meet certain assumptions to be valid. A table showing how Difference-in-Differences is calculated is here.

For the theory of change itself, start with a graphic that illustrates the expected cause and effect between activities, intermediate results, and longer-term outcomes. Then add a narrative spelling out the detailed cause and effect as well as the underlying assumptions. For the graphic, work backwards from the ultimate outcomes or impacts; this is called "backwards mapping" and makes it is easier to chart the logic.

Examples of water fund theories of change can be found here and here.

Choosing an Evaluation Technique

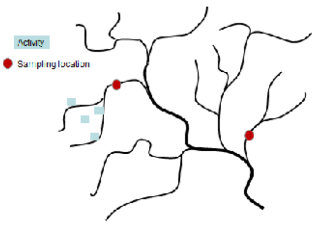
Choosing an evaluation technique for a water fund warrants careful consideration. Monitoring and evaluation lethal mistake #2 is vague objectives and/or poor experimental design. Avoid this mistake by making an informed choice.

First, decide on contribution or attribution. "Contribution" means a water fund contributes to changes but may not be the only source of the changes. Contribution involves measuring indicators before and after water fund activities. This only shows the project's contribution to changes because other external factors could have contributed to the changes as well. For instance, there may have been local projects working on similar activities as the water fund or local rains may have been favorable. (See *Appendix 4: Statistical Considerations* for more on this topic.)

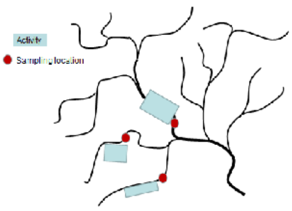
"Attribution" means that changes can to attributed to water fund activities. This is done via an estimation of what would have happened anyway had there been no water fund activities — known as the "counterfactual" (for more on this concept, see here). For water funds, the counterfactual is usually a control/comparison area that is similar to the water fund area except it has no water fund activities.

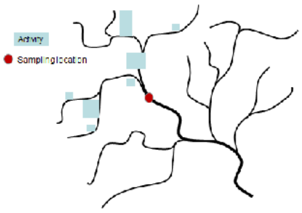
"Contribution" evaluation techniques should only be chosen if the water fund is implementing activities that have already been rigorously measured with control/comparison sites in similar contexts, and there is no need to repeat these measurements. All other water funds should use an evaluation technique that allows for attribution (i.e., with control/comparison sites), so we can build the evidence base for water funds.

The six options for evaluation techniques are in the table below.[3] The options are presented from most to least rigorous. (Choosing more than one option is possible).

Option	*Advantages*	*Limitations*
Paired watersheds  	• A watershed with water fund activities is matched with a control watershed that is similar but has no activities. • This option allows for attribution. • Works well on headwaters of watersheds.	• Depends on accurate matching of paired watersheds. • Need two or more years of baseline data to show pre-existing differences (if any) between paired watersheds. • Fires, insect infestations, water abstraction, construction, road building, or any large-scale land-

[3] Pages 47-59 of the Water Fund Monitoring Primer here have more about each option.

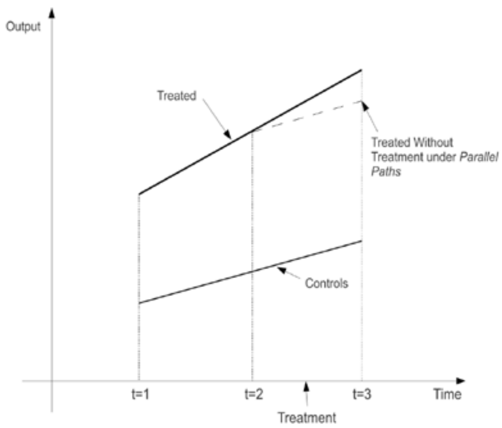
Option	*Advantages*	*Limitations*
	• Well suited for micro-watersheds.	use changes in either watershed can negate the comparison.
Multiple watersheds 	• This compares three matched watersheds: water fund, control, and intact reference sites. • This option allows for attribution. • Shows if the water fund watershed is diverging from the control watershed and converging with the reference watershed that is in the desired condition.	• Requires an intact reference site for comparison. • The three watersheds have to respond to inputs in similar ways. • Need two or more years of baseline data to show pre-existing differences among watersheds. • Fires, insect infestations, water abstraction, construction, road building, or any large-scale land-use changes in a watershed can negate the comparison.
Above/below multiple sites 	• Measures difference between water parameters upstream and downstream of water fund activities. • Change can be measured in a shorter period of time than at a watershed scale. • Works well for a short reach of river. • Results can be used to show the benefits of activities that are being implemented throughout a larger watershed.	• Attribution depends on the condition and context of the upstream sites, and if highly degraded, then it may mask benefits from water fund activities. • Need two or more years of baseline data to show pre-existing differences between upstream and downstream sampling points. • Other streams or rivers joining between upstream and downstream sampling points can compromise the comparison.
Monitoring downstream of multiple sites 	• This measures changes over time at several locations downstream of water fund activities. • Provides multiple sample points for comparing before/after changes.	• Does not allow for attribution. • Assumes there are no changes upstream of the treatment area that could influence the results. • Watersheds may respond differently to the same inputs.

Option	*Advantages*	*Limitations*
Two watersheds without baseline data 	• This compares a watershed with water fund activities and an unmatched control watershed. • Does not require a baseline calibration period.	• Measured difference may be due to inherent differences in the watersheds. • Does not allow for attribution. • Not known if the two watersheds started with the same conditions or respond differently to rainfall or water fund activities.
Monitoring downstream at a single site 	• This measures changes over time at a single downstream location. • Often seen at existing monitoring site with a long record of data collection. • Useful in monitoring long-term trends in a large watershed. • Contributes data that can corroborate other data sources.	• Does not allow for attribution. • Shows long-term trends but not if water fund activities changed these trends.

Source: *Primer for Monitoring Water Funds*

For water funds working in headwater areas, a “paired watersheds” design works well if micro-watersheds — very small watersheds in which water fund activities can cover the majority of the watershed — are chosen that are representative of the larger water fund area. A “multiple watershed” design can also work if one of the sites is an intact reference watershed, another is the control watershed, and the third is the water fund watershed. Both of these designs allow for attribution of water fund activities to the changes observed because they include matched control sites. Finding matched water fund and control sites is not easy (text box).

The challenge of finding matched control sites
Control sites need to be similar to water fund activity sites. To decide this, watersheds are often matched on precipitation, dominant land use, soil type, slope, aspect, and distance to the nearest road. No control site is a perfect match, however. The most rigorous test of matched sites is known as the ‘parallel paths assumption’. Here the comparison sites do not have to be identical, but they do need to have parallel trendlines before water fund activities (“treatment”) began.



Source: Mora & Roggio 2012

Once an evaluation technique is selected, one or more key evaluation questions should be chosen. These are the questions that the evaluation is designed to answer. For example, did the water fund's upstream water conservation activities increase dry-season water flows downstream, or did restoring local flora upstream reduced sediment loads downstream, and if so, by how much?

Selecting Indicators

Choosing indicators can be a daunting task for people new to monitoring and evaluation but is not that difficult. There are two primary kinds of indicators in a project: output and outcome.

Output indicators track key water fund deliverables. Examples are # of hectares protected, # of local beneficiaries, and # of trees planted.

Outcome indicators track the big things the water fund wants to achieve in the short to medium term. Examples of water fund outcome indicators include a 10% reduction in turbidity between the baseline and five years later for the same high-flow levels, and metric tons of CO_2e emission avoided between the baseline and five years later.[4] Decide on outcome indicators before selecting output indicators.

Choose outcome indicators carefully because the indicators chosen become how a water fund defines success. Get consensus among key water fund stakeholders on the outcome indicators to ensure there is a shared understanding of what success means.

Four categories of outcome indicators are important for the sustainability of a water fund: (i) water quality and quantity; (ii) biodiversity/landscape scope or quality; (iii) financial; and (iv) human well-being. All categories of indicators may not be relevant to all water funds, however. For each relevant category, select at least two outcome indicators. This guards against problems with one indicator compromising a project's evaluation.

When selecting indicators, it is better to use existing indicators than create new ones. This helps ensure the indicators are credible, valid, and feasible and allows for reliable comparisons across time, place, and measurers. Common water quality indicators can be found here, and widely used people indicators can be found here.

When selecting indicators, ask three questions about each indicator:

- What action would be taken based on the indicator results?
- Are there the resources necessary to implement the action?
- Are managers committed to taking the action if needed?

Answering these questions helps avoid lethal mistake #3: Failure to act on and not just review the data collected. In short, know how each indicator will be used before including the indicator in the project's monitoring and evaluation plan.

[4] Outcome indicator should be SMART: specific, measurable, achievable, realistic, and time-bound.

The exception to this is socioeconomic indicators. Oftentimes the full range of socioeconomic changes due to water fund activities are unknown. Here it is helpful to include indicators in a household survey to measure all the elements of human well-being. These socioeconomic indicators are usually only measured at baseline and endline.

Monitoring and Evaluation Plan

Defining the desired changes, developing a theory of change, choosing an evaluation technique, and selecting indicators are the building blocks for drafting a monitoring and evaluation plan. This plan documents the choices made and is updated periodically as the water fund learns what works and what does not.

For the monitoring and evaluation plan, include only a small number of output and outcome indicators — 10 to 20 is a reasonable range in our experience. A new water fund may have more output indicators that outcome indicators, and a mature water fund with few field activities may have more outcome than output indicators. But limit the number to only the indicators that will be actually used by managers.

Finally, each indicator in the monitoring and evaluation plan needs to be precisely defined so it can be measured in a replicable way. There are two ways to do this: in a table like the one here or with indicator reference sheets like the ones here.

A draft outline of a monitoring and evaluation plan is in Appendix 3. An example of a water fund monitoring and evaluation plan can be found here.

Data Collection

At a minimum, all water funds should collect data on key water parameters. A water fund that collects no water-related data will struggle to justify its existence. A water fund also needs to collect information on local people's perceptions of the water fund activities as part of the minimum monitoring and evaluation activities or risk unpleasant surprises (text box).

The starting place for data collection should always be a review of existing data sources. Build on what already exists whenever possible. Only if the existing data sources are inadequate should a water fund undertake its own data collection.

Data collection is a specialized skill, so use trained people to collect data. Bad data are worse than no data because they are misleading. Partnering with an organization that specializes in data collection or partnering with academics can work well.

The frequency of data collection depends on the indicator and may range from every 30 minutes for water monitoring stations to once a year for local people's perceptions about the water fund.

All implementation data should be 'geo-tagged' with elevation, latitude, and longitude. Knowing where water fund activities took place is vital for understanding cause and effect at the local level.

Monitoring the Achilles Heel of Water Funds

If there is a single most vulnerable spot for water funds, it is participation by local people. People upstream can choose to join water fund activities or not — it's voluntary. Our working hypothesis is that most people choose based on the economic benefits. But the economic benefits can change. If, for example, the sale price of a crop increases quickly, and farmers can earn more by expanding into the riparian buffer strips that the project helped protect, farmers may do so. Or if a dairy farmer wants to increase production, he can choose to take down his fence protecting a riparian area and graze the buffer strip. In both cases, the local river water quality is likely to decline because of the change. Participation in watershed protection activities needs a critical mass of providers to achieve the benefits. If the number of people who participate drops precipitously, watershed protection can be compromised. Thus, include in the monitoring and evaluation plan at least one indicator on people's perceptions. An example of this is % of people who say they support water fund activities. Use the data collected to highlight problematic sub-watersheds where participation is low or ineffective. Only by monitoring local support for the water fund activities can we know if support for the activities is waxing or waning. If support is dropping, understanding why becomes critical for water fund success. This monitoring can be done cheaply via focus group discussion or mobile phone SMS surveys.

For evaluations, there are two critical times for data collection in a project: baseline and endline.[5] The baseline is a point in time against which changes can be measured and compared. Baselines are required for evidence-based conservation. Usually baselines are done before a project starts, but baselines can be conducted after water fund activities begin as well. The consequence of a late baseline is that early water fund gains may not be captured because they have happened already. A late baseline is better than no baseline if most of the project activities have not happened yet.

[5] Projects may also have formative evaluations or mid-term evaluations to guide the adaptive management of project activities.

Deciding whether or not to include a socioeconomic survey at baseline and endline is a key decision (text box).

Baselines water parameters for a water fund can take several years to collect. The year-to-year variations in a watershed can be substantial, and only with multiple years of data can one see if water fund and control sites respond in similar ways to precipitation because of the noise inherent in the data. Starting activities in parts of the project area away from the chosen treatment and control sites is one way to not delay water fund startup until multi-year "calibration" baselines are completed.

Data collection for an evaluation should take place approximately every five years (depending on the outcome indicators) or at the end of a donor funding cycle, whichever comes first.

When thinking about data collection, it is helpful to define each stakeholder's role. One way to do this is via a "bullseye" summary of stakeholders' primary monitoring and evaluation roles like the one shown here.

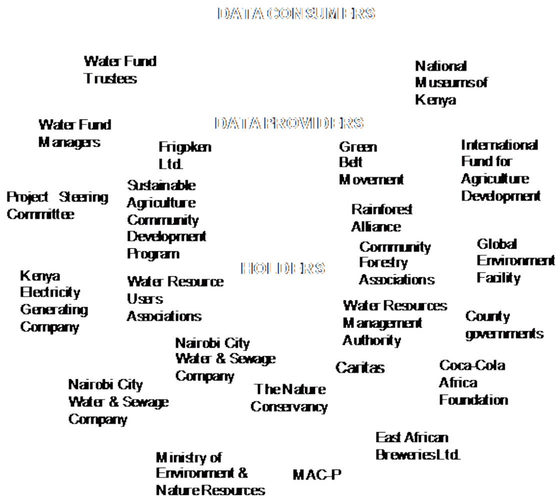
Socioeconomic surveys in a water fund

Socioeconomic surveys can be powerful tools for generating political and donor support for a water fund but are expensive. Water funds that aim to improve local people lives should include some form of socioeconomic data collection. If there are less than 100 households likely to benefit from water fund activities, focus group discussion to learn what local people perceive as the costs and benefits of water fund activities may be sufficient. It usually takes from three to six focus groups to elicit 90% of local themes. The cost is a few days of time and about US$5,000 for travel, room and board, two facilitators, snacks, and a summary report at the end. If the number of households is between 100 and 300, interviews should be undertaken for all households (a comprehensive survey). If there are more than 300 household, a randomized sample of the household should be conducted. Randomized household surveys cost between US$60,000 and US$100,000 for transport, meals, local enumerators, data collection, data analysis, and report writing. It takes about six months from design to final report. For a randomized household survey, a sample size of 1,000 households is often a good tradeoff between ability to detect change and cost. (For more on sample size and statistical considerations, see Appendix 4.)

Ethical Issues

Data collection and analysis need to fully align with The Nature Conservancy (TNC) core values, especially integrity beyond reproach and respect for people, communities, and cultures.

This means familiarizing one's self with the ethical issues involved in collecting data from people or conducting research on human subjects. The best starting place for this is a free online course here. Note that any plan by TNC staff to collect data from people needs to be approved by the Chief Scientist or his delegate.[6]



Source: UTNWF M&E Plan

[6] The details and the forms can be found on Connect here. It usually takes less than two days to get approval. Note that all data collection from people should start with the Free, Prior, Informed Consent (FPIC) of the respondent. An example of an approved FPIC statement to be read to the potential respondent can be found here.

When selecting controls for people indicators, there is an ethical problem with not providing the same benefits to people in the controls as those in the water fund or treatment areas. One way to address this issue is to sequence the implementation using a stepped-wedge design. This is done by rolling out an activity in stages, making it possible to randomly select the order in which the participants receive the activity so that by the end of the project, both treatment and control groups have benefited from the activity.

Data Analysis

The most frequent failure point for monitoring and evaluation is data analysis. This is lethal mistake #4: Failure to turn data into information. Data tend to sit on hard drives because the analysis is a big task, and the data pile up until the task becomes overwhelming. Another dead dataset is the result.

The key to avoiding this pitfall is to start the data analysis early to identify and correct problems with sample design and sample collection. Then, automate the analysis task to the extent possible. Technology is the data analyst's friend. There are now tools for big data analysis including Microsoft Power BI, DHIS2 and Google Earth Engine. Whenever possible, use software to automate the data analysis so that it can produce summary reports or a data dashboard simply by inputting the data collected, or in the case of Google Earth Engine, running an algorithm. Pilot it, fine tune it, and then automate the data analysis to feed into a report or data dashboard. You'll thank yourself later (and then us we hope).

Time lags between activities and measurable impacts on water quality
Be forewarned that the larger the watershed, the longer it takes for water fund activities to have a measurable change on water quality. A number of factors influence lag time include the initial state of degradation of the landscape, hydrology of a watershed, vegetation growth, transport rate and path, and pollutant properties. The time lag is highly site and pollutant specific and can range from months to years.

Data Quality

Reliable and valid data are critical to the long-term success of a water fund (to learn what constitutes reliable and valid data, see here).
High-quality data starts with quality assurance procedures. These are usually written protocols to standardize data collection so that different people collect the data the same way. Examples of water-related data collection protocols can be found here and here. Assessing data quality is also important. This usually includes randomly selected spot checks to verify the reported data and follows data quality assessment checklists like the ones here and here. Data storage also matters (text box).

Data storage
For online storage for water fund data, the River Basin Information System (RBIS) is a data repository tailored to hydro data that TNC helped develop. It is free and comparatively easy to use and is the recommended data storage tool for all water funds.

Sharing the Learning

Given the number of water funds globally, substantial learning is no doubt happening at the local level. A monitoring and evaluation system can help capture this learning. Sharing this learning internally and externally with water fund supporter helps future projects and future stages of current project.

Include resources to document and share lesson learned in the monitoring and evaluation budget. Many water funds, however, have no monitoring and evaluation budget. We address this issue and another common constraint below.

Addressing the Two Biggest Constraints

Two frequently cited reasons for water funds not having monitoring and evaluation systems are insufficient staff capacity and too little funding.[7] Here's advice on both.

There are two primary ways to address staff capacity issues: building it or buying it.

Building it means hiring staff such as a monitoring and evaluation specialist. The monitoring and evaluation specialist oversees the data collection, does the data analysis, and drafts reports for water fund managers and donors. This is the person who turns data into information.

Buying it means hiring outside experts to do the water fund monitoring and evaluation. This is usually split between one team that does the monitoring and another team that does the evaluation (both are specialized skills).

Addressing too little funding is more challenging. The best approach is to build monitoring and evaluation costs into the water fund budget during the design stage. But how much is enough? The short answer is about 10% of total project costs should be earmarked for monitoring and evaluation. The longer and more accurate answer can be found here.

Every water fund needs to monitor at least one key water quality or quantity parameter (to know if it is making a difference) and local people's perception of water fund activities (to ensure local people continue to support water fund activities). This is the absolute minimum a water fund should do. The practical minimum, however, should be determined by the number of outcomes the water fund is trying to achieve.

One way to allocate scarce funding is with a hierarchy of monitoring and evaluation needs. Below is a list of monitoring and evaluation activities in rough order of importance to a 'typical' water fund that is in the startup phase. If resources are scarce, allocate monitoring and evaluation funding starting with the first on the list and working down as resources allow. (NB: the cost estimates provided are indicative and will vary from context to context.)

1. *Implementation monitoring*. Implementation monitoring is the top funding priority because if implementation of water fund activities is done poorly, the water fund is likely to fail. Implementation monitoring costs from US$18,000 to US$70,000 per year. At the low end, if we assume a water fund staff working 25%

[7] Another primary reason is a lack of incentives. Even when people have sufficient capacity and funding, they are often reticent to develop and implement monitoring and evaluation because they do not appreciate its potential added value, and instead perceive it as an accountability hope and hinderance.

at a grade 7 mid-point salary of a middle-income country, then this costs about US$15,000 a year. If we assume data are collected for three to five outputs (deliverables) that are all relatively easy to monitor such as # of meters of riparian areas fenced, # of trees planted, or # of hectares with invasive plants removed, then data collection tools, transport, and supplies might cost about US$2,000 a year. Adding in perception monitoring of local resident via annual focus group discussions would add about US$1,000 a year. Thus, the minimum cost for implementation monitoring is US$18,000 a year. For projects that need more than basic implementation monitoring, consider hiring a monitoring and evaluation specialist, using monitoring and evaluation software such as DHIS2 for the project, collecting implementation data via mobile phone (part of DHIS2), and using mobile phone SMS to measure local people's perceptions of water fund activities. This is a stronger form of implementation monitoring, and for the Nairobi water fund, the costs to do this were US$50,000 in one-time costs and US$70,000 in annual costs (mainly for the M&E specialist's salary and benefits). These costs were covered by the 10% set aside for monitoring and evaluation in the project budget.

2. *Setting water quality and/or quantity baselines.* This is the second priority because establishing a baseline for the hydrological behavior of a watershed can take several years because of the inherent noise in the data. These baselines need to ideally start several years before water fund activities begin in a location. Consider starting the baselines as part of the pilot phase or design phase of a water fund. Costs for these baselines are covered in the specific types of monitoring below.

3. *Weather monitoring.* This measures the physical characteristics of the watershed such as precipitation, temperature, humidity, and wind. At a minimum, a water fund should measure precipitation daily in at least two places inside the target watersheds because of the relationship between local precipitation and water quality and quantity. The simplest solution is to install rain gauges (US$20-US$50) that are read and recorded daily. More sophisticated solutions include automated tipping bucket gauges with dataloggers that cost US$350 to US$800 or fully integrated weather stations that cost US$500 to US$2,000.

4. *Water-level monitoring.* This measures water level that is then used to calculate flow based on rating curves developed for the specific river or stream and can be used to measure changes in water flows due to water fund activities. Water-level staff gauges and local gauge readers are the least-cost option for water-level monitoring and can help locals to become interested in and part of the water fund monitoring. This is likely to cost several hundred dollars for each staff gauge and installation. The more reliable system is to install automated water monitoring stations with water level loggers like the ones here, here, and here. In the Nairobi water fund, there are 26 water-level monitoring stations and one person who collects the data and does the data logger maintenance every few months. The

total for this was US$26,000 in one-time costs and US$650 in annual costs. Such costs, however, vary based upon the specific monitoring system and local economic factors. In Rio de Janeiro, routine monitoring includes climate, discharge, and turbidity. Annual costs range from US$10,000–15,000 for contracts, equipment upkeep, data analysis, and sample collection.

5. *Suspended solids monitoring*. Reducing suspended solids is often a priority for water funds because these reduce water treatment costs and reflect reduced soil loss and ground cover (habitat) disturbance. Suspended solids monitoring should be done in conjunction with collecting data on local precipitation and flow because of the relationship of rain intensity and runoff to suspended solids. Monitoring can be as simple as measuring turbidity if a relationship between total suspended solids and turbidity can be developed. A portable turbidity meter that costs about US$1,300, downloading synchronous local government rainfall data, and someone to do periodic data collection in the field is the minimum needed for suspended solids monitoring. A more robust monitoring system involves grab samples from local water courses with total suspended solids measured from the samples in a lab. In the Nairobi water fund, the cost to install two automated water-level stage and turbidity monitoring stations and have a part-time person do the routine maintenance was approximately US$17,000 in one-time costs and US$650 a year.

6. *Remote sensing imagery*. This is a cost-effective tool to measure land-use and land-cover changes due to water fund activities. Satellites imagery can be used to measure changes in forest or plant cover over large areas. The cost is usually several days of a GIS expert's time to download the images and analyze them for a watershed. For smaller areas such as a micro-watershed, aerial imagery from a drone or light plane can provide the high-resolution imagery needed to track small changes on the ground. These costs can range from a few thousand dollars to much more, depending on area covered and resolution required.

7. *Bioindicator assessment*. Because aquatic insects are constantly exposed to water quality and quantity stressors, their presence, absence, and relative abundance make them good indicators of aquatic conditions. The larger insects that can be directly observed (called "macroinvertebrates") are practical, inexpensive, and effective pollution monitoring devices. Surveys of animals in the water fund area can also be used to identify changes over time. Macroinvertebrates provide short-term indicators of environmental changes, while fish distribution and abundance provide longer-term indicators. The Rio de Janeiro water fund paid US$8,000 for a fish survey at 14 sites that will be redone every five years and US$15,000 for a bird survey at 36 sites that will be redone every three years. The Upper Tana-Nairobi water fund paid US$24,000 for invertebrate and bird surveys in 11 sites that will be redone after four years.

8. *Socioeconomic surveys*. These are baselines, follow-on, or endline socioeconomic surveys to measure changes in local people's lives from the water fund using

treatment and control groups. Such surveys quantify the benefits to local people from water fund activities. The 1,000-household randomized survey of the Nairobi water fund cost US$71,656 including the analysis and report writing. Socioeconomic surveys can be powerful tools for generating political and donor support for a water fund but are expensive.

Conclusion

Water funds are a compelling idea. It is why there are more than 35 of them today. Yet there is little evidence of their people and nature benefits to date. This is not because the benefits are not there. Instead, it is because we are not measuring these benefits and because we are not maximizing these benefits by learning what works and what does not. We are selling water funds on the strength of the idea and not the evidence that the idea works. Over time, this will become increasingly tenuous and could expose the organization to reputational risk.

To sustain the growth in water funds, we need evidence of impacts, and to do this, we first need a strong focus on good implementation and then an equally strong focus on impact evaluations. In short, we need to recapitalize the water fund idea by adding evidence, and for this we need monitoring and evaluation for water funds.

Useful Links (curated)

Water quality

Water quality measurements of interest for water funds
Water quality monitoring and assessment
National field manual for the collection of water-quality data
Monitoring water quality
Monitoring and evaluating nonpoint source watershed projects

Water quantity

Surface water flow measurement for water quality monitoring projects
How streamflow is measured
Discharge measurements at gaging stations
Weirs - open channel flow rate measurement

Weather data

Technical manuals World Metrological Organization e-library
National Weather Service training for new station installation

Biodiversity

Rapid biological assessment protocols: an introduction
Rapid bioassessment protocols
Stream corridor restoration handbook
Example of a biodiversity assessment in a water fund (Kenya)
Example of an ecological integrity study in a water fund (Ecuador) (Spanish)

Socioeconomic

[Detailed design of a household survey]
[Terms of reference for a baseline household survey]
[Example of a baseline report]

Appendix 1: Checklist for Monitoring and Evaluation of a Water Fund

- ☐ Secure funding for a water fund monitoring and evaluation system.
- ☐ Define desired changes due to the water fund.
- ☐ Get consensus among partner organizations on desired changes to be addressed.
- ☐ Define approximate boundaries of water fund and priority areas.
- ☐ Draft theory of change.
- ☐ Choose an evaluation technique.
- ☐ Select outcome indicators and ensure they are SMART.
- ☐ Select output indicators to monitor implementation.
- ☐ Get agreement with partner organizations on outcome indicators.
- ☐ Write a monitoring and evaluation plan for the water fund.
- ☐ Build or buy the needed capacity to implement the monitoring and evaluation plan.
- ☐ Collection baseline data.
- ☐ Do baseline data analysis.
- ☐ Implement data collection protocols for quality assurance.
- ☐ Begin implementation monitoring data collection.
- ☐ Automate implementation monitoring data analysis.
- ☐ Begin using monitoring data to adaptively manage water fund activities.
- ☐ Do periodic data quality checks.
- ☐ Conduct follow-on data collection for outcome indicators.
- ☐ Analyze and document what worked and what did not (lesson learned).
- ☐ Share learning internally and externally.

Appendix 2: Glossary of Key Terms

Activities – Things that a person or group do to implement a water fund.

Adaptive management – This is an iterative process for continually improving project management by supplying timely information to project managers on potential issues that may require changes in the current approach.

Baseline – A point of reference in time against which changes can be measured over time. A baseline study describes the initial conditions and indicator levels of a project.

Control – A comparison site or group that is similar statistically to the participant (treatment) site or group but is unaffected by the project activities (or treatment).

Counterfactual – Shows what would have happened had there been no project activities. Usually comprises a control site or group.

Endline – Measurement made at the completion of a project to compare the conditions after the project with the baseline conditions.

Future – That period of time in which our affairs prosper, our friends are true, and our happiness is assured.*

Indicator – A trend or fact that indicates the state or level of something. Synonym for 'measure'. A good indicator is measurable, precise, consistent, and sensitive to change.

Impact – The positive or negative, primary or secondary long-term effects produced by a project, directly or indirectly, intended or intended. An impact is longer term than an outcome.

Implementation – This is the process of putting plans into action.

Input – The staff time, contracts, equipment, travel, etc. invested in an activity.

Measure – A standard unit used to express the size, amount, or degree of something. Within monitoring and evaluation, it is synonymous with indicator.

Outcome – A specific, measurable, achievable, realistic, and time-bound (SMART) result that leads to the impact. Similar to an intermediate result but measurable and time-bound. More specific and shorter term than an impact.

Output – A deliverable over which the project implementers have control.

Reliability – If one tests and then retests are the results the same?

Resolution – The precision of a spatial or temporal measurement.

Treatment – A site or group that is directly affected by project activities.

Validity – Does an indicator measure what it purports to measure?

*Ambrose Bierce, 1906

Appendix 3: Outline for a Monitoring and Evaluation Plan

Introduction

Theory of Change

Narrative and graphic theory of change

Monitoring

- Monitoring strategy
- Output indicators
- Data sources and methods
- Data analysis frequency
- Data quality protocols

Evaluation

- Evaluation strategy
- Outcome indicators
- Data sources and methods
- Baselines

Learning

Responsibilities

Timeline

Budget

Appendix 1: Indicator reference sheets

Appendix 4: Statistical Considerations

In all watershed projects, including water funds, the basic premise behind monitoring is to provide evidence of changes resulting from project interventions. The ability to demonstrate such changes depends not only the efficacy of the interventions themselves, but also on the experimental design and analytical approaches used to evaluate these changes. While only some readers of this document are likely to plan and perform data analyses themselves, all readers can benefit from familiarity with key considerations related to statistical analysis.

Explanatory Variables

Environmental systems are complex. For any given parameter or variable, a large number of environmental processes may all contribute to the observed effect. For example, while Total Suspended Solids (TSS) measurements may reflect changes due to implementation of agricultural best management practices (BMPs), they may also reflect seasonal or annual changes in precipitation and streamflow, differences in soil conditions, or other sources of sediment such as streambank erosion. In this example, TSS is the *dependent variable*—it is the variable being influenced by other factors. The factors or variables that explain the changes in TSS are known as *explanatory (or independent) variables.* While BMP implementation may be the primary explanatory variable of interest, there are often many other factors that also influence observed TSS values.

Key takeaway: In most cases, the effect you want to measure is influenced by factors beyond just your intervention alone. Take action to account for these 'confounding variables' by implementing good experimental design and monitoring of the most important explanatory variables.

A major statistical consideration for monitoring concerns addressing the influence of these other explanatory variables. If these other explanatory variables are not considered, it can be impossible to distinguish changes resulting from project interventions versus those changes that result from other activities or processes.

Selection of experimental design is the most important means to address the confounding impacts of other explanatory variables (see *Choosing an evaluation technique*). Where conditions permit, paired sampling from paired watershed and above/below-before/after designs can effectively control for such confounding variables. In such cases, variables measured in the control or upstream watersheds can themselves be considered explanatory variables.

But beyond control variables, there are a number of common explanatory variables that should be considered for inclusion within monitoring plans. The determination of which variables should be monitored is project-specific, but at a minimum, water fund managers should ensure that such variables have at least been considered. Common categories of explanatory variables are listed below. For any one category, there may be multiple variables that could be relevant for a given project.

- Land use and land management changes
- Seasonality or other cyclical patterns

- Hydrologic and meteorological variables (e.g., discharge, precipitation, air temperature)

Statistical Analysis

While the decision about which specific statistical model to use is beyond the scope of this document, there are general concepts that are useful to keep in mind. Statistical testing is the application of probability to make inferences from data. At its most fundamental, statistical testing allows us to describe—in quantitative terms—how likely it is that any observed differences are due to chance alone. In the case of watershed monitoring, statistical testing is particularly important due to the large spatial and temporal variability of many environmental parameters.

> **Key takeaway:** Statistical analysis is critical during *both* pre-intervention planning and post-intervention evaluation. Monitoring results without appropriate statistical testing lack credibility.

At the outset of a project, statistical analysis of pre-intervention conditions allows for appraisal of the *statistical power* of a given monitoring approach. For example, minimum detectable change analysis can provide an estimate of the expected percent change that might be observable—given historic watershed conditions, selected experimental design, and sampling frequency and duration. In many cases, project supporters may be surprised at the size of this minimum detectable change, which often can be much larger than the actual anticipated impact from interventions. Such statistical analysis can provide valuable, early information to adjust monitoring plans and set expectations.

After interventions have been implemented, statistical analysis allows for inferences of any observed trends. In cases where interventions are completed in a relatively short period of time, this could be considered a "step change"—with discrete pre- and post-intervention periods. In cases where interventions are implemented gradually over time, this suggests a linear or monotonic type of trend. Different statistical models are applicable to each of these different trend types. Thus, understanding how a project will be implemented is important for choosing among statistical tests.

The previous section highlighted the importance of good experimental design and identification of explanatory variables beyond the primary (treatment) variable of interest. But good experimental design and monitoring of explanatory variables can only go so far in addressing observed variations in water measurements. Not all sources of variation can be controlled or adequately monitored and other sources of error exist. Statistical analysis can then be used to describe the resulting uncertainty. In order for watershed monitoring findings to hold any credibility, such statistical analysis is imperative.

Importantly, while statistical analysis can provide information about the *statistical* significance of results, this should not be confused with *biological* or *social* significance. Even where a statistically significant change might be observed, there may be no practical "real world" significance.

Sampling Frequency and Duration

The number of samples collected will affect the statistical power of a given monitoring approach. In general, increasing the frequency and/or duration of sampling will improve (decrease) the minimum change that can be detected. However, there are limits to what can be gained. As sampling intervals become smaller (e.g., less than weekly), fewer 'new' information is provided by each sample. This is because samples within a small window of time often exhibit correlation (due to the temporal proximity). Statistical tests can evaluate and adjust for such *autocorrelation*. In contrast, increasing the duration of the sampling period avoids this challenge but at the obvious expense of increased delays, costs, etc.

Beyond these generalizations, it difficult to make prescriptive recommendations regarding sampling frequency and duration. Specific choices depend upon the variable being monitored, conditions within the waterbody of interest, budget, and other constraints. The critical point to keep in mind is that these choices will have implications for subsequent statistical analyses.

References and Further Reading

J. Higgins, A. Zimmerling, "A Primer for Monitoring Water Funds", The Nature Conservancy, 2013: https://www.nature.org/media/freshwater/Water_Funds_Primer_on_Monitoring_2013.pdf

US EPA, Nonpoint Source Monitoring: TechNOTES. US EPA (2015): https://www.epa.gov/nps/nonpoint-source-monitoring-technotes