

GEF Evaluation Office

Foundations of Success

GEF IMPACT EVALUATION

Final Report on a Proposed Approach to GEF Impact Evaluation

Impact Evaluation Information Document No. 2

Prepared by Foundations of Success

September 2007



This paper was commissioned by the GEF Evaluation Office (GEF EO) as an input into its program of **Impact Evaluation**.

A first annual report on this program will be presented to the GEF Council at its November 2007 meeting. The findings, interpretations, and conclusions expressed herein are those of the authors and do not necessarily represent the views of GEF Evaluation Office, the GEF Council, or the Governments they represent. The authors of this document would welcome any comments or suggestions on its contents.

The papers in the Impact Evaluation information document series, as of September 2007, are:

1. Approach Paper to GEF Impact Evaluation – *Brann and Todd*
2. Final Report on Proposed Approach to GEF Impact Evaluation - *Foundations of Success*
3. GEF Biodiversity Policy Review - *Foundations of Success*
4. Methodological Challenges in Impact Evaluation: The Case of the Global Environment Facility – Todd and Vaessen
5. Priorities and indicators for Global Environment Benefits from Biodiversity: The current international architecture – *Nair*
6. Case Study Methodology – *Conservation Development Centre*
7. Case Study: Bwindi Impenetrable National Park and Mgahinga Gorilla National Park Conservation Project - *Conservation Development Centre*
8. Case Study: Lewa Wildlife Conservancy – *Conservation Development Centre*
9. Case Study: Reducing Biodiversity Loss at Cross-Border Sites in East Africa *Conservation Development Centre*
10. Impacts of Creation and Implementation of National Parks and of Support to Batwa on their Livelihoods, Well-Being and Use of Forest Products – *Namara*
11. Protected Areas and Avoided Deforestation: A Statistical Evaluation – *Andam, Ferraro, Pfaff and Sanchez-Azofeifa*

Global Environment Facility

Director of the GEF Evaluation Office: Robert D. van den Berg

Impact Evaluation Team

Task Manager: David Todd, Senior Evaluation Officer

Evaluation Analyst: Divya Nair, Junior Evaluation Professional

Co-reader: Lee A. Risby, Evaluation Officer

Contents:

Executive Summary	i
Background	1
Purpose of This Work	1
Modes of Impact Evaluation	1
Evaluation State-of-the-Art	4
Types of Theories of Change Tools	8
Methods	10
Sample	11
Determining Utility of Approach to Evaluate	11
Determining Utility of Approach to Aggregate Results	11
Findings	13
Analysis of Pre-GEF-3 Projects	13
Analysis of GEF-3+ Projects	16
Applying a Theory of Change Approach to the Sample	17
Using a Theory of Change Approach to Aggregate Project Results	23
Recommendations	27
Vision for an Ongoing Evaluation Program Based on a Theory of Change Approach	27
Necessary Conditions to Undertake a Theory of Change Evaluation Approach	34
Conclusions	36
Piloting the Theory of Change Evaluation Approach Draft Terms of Reference	40
1. Background and Rationale	40
2. Purpose	40
3. Scope	42
4. Methodology	45
5. Pilot Workplan	49
6. Key Audience and Partners Involved	51
Attachment 1	53
Attachment 2	55
Attachment 3	i

Executive Summary

The purpose of this project was to develop an impact evaluation approach for GEF projects. This report is designed to serve as a proposal for such an evaluation approach based on the “Theory of Change” concept.

We reviewed twenty projects – ten pre-GEF-3 projects and ten GEF-3 projects – in order to test the potential of a Theory of Change approach. Based on our analysis, we conclude that a Theory of Change approach to impact evaluation is, indeed, warranted. This approach provides a roadmap for analyzing how and why an initiative works. It delineates the pathway of an initiative by making explicit both the expected results of the initiative (early, intermediate, and long term) and the actions or strategies that will lead to the achievement of results.

In the sample we reviewed for this work, we found that there is fairly stark difference in terms of logic and clarity between pre-GEF-3 projects and GEF-3 projects with the newer projects significantly better articulated than the old ones. Of the older GEF projects, however, about half include enough desired project results/outcomes to be able to construct at least logic models and simple results chains.

In order to accommodate the diversity of pre-GEF-3 and GEF-3+ projects, we have found that there must be at least two levels of the Theory of Change approach. The first approach must be able to accommodate those projects for which the Theory of Change is not very explicit, assumptions are not articulated, key variables for impact evaluation are not identified, and data on key indicators are not collected. In order to evaluate projects under these conditions, we recommend taking a Theory of Change approach that is represented by *logic models*.

For those GEF projects that do have well articulated interventions and expected results – and the assumptions that link the two – we advocate a more rigorous form of the Theory of Change approach. This approach is best represented by *results chains*. Applying the more rigorous results chain approach allows evaluators to determine the extent to which projects have been successful. In particular it facilitates analysis of not only the extent to which a given intervention worked, but also why it worked or did not work.

Adopting a Theory of Change approach will provide GEF with a mechanism to look at its results across its portfolio of projects instead of looking at results exclusively in terms of individual projects. It will help GEF better understand what has worked and what has not worked in the past. And finally, a Theory of Change approach to evaluation will facilitate GEF’s efforts to enhance its impacts by documenting results, accelerating learning, and providing a mechanism for systematic adaptation.

Background

The GEF Evaluation Office (EO) is seeking an effective approach to measure the impacts of GEF-funded activities. This report summarizes the results of a review of a sample of projects and provides recommendations for designing an evaluation approach based on the Theory of Change model.

Purpose of This Work

The purpose of this study was to develop an impact evaluation approach targeted to GEF's biodiversity focal area that should – with some modifications – be applicable to other GEF focal areas. There are three main components of this work.

1. A proposal for an evaluation approach based on a “Theory of Change” concept;
2. Analysis of policy decisions affecting GEF programmatic priorities and M&E work; and
3. Analysis of how a GEF-wide evaluation approach might link to regional and global indicators.

This report represents the first component listed above. Specifically for this aspect of the work, we aimed to:

- Identify change strategies underlying a sample of projects and determine the prevalence of these strategies in the biodiversity portfolio;
- Construct theories of change using assumptions and variables identified in project documents;
- Determine the feasibility of aggregating results across projects and possibly linking them to global indicator databases; and
- Determine if a common impact evaluation approach could be proposed for all GEF projects or if approaches need to be tailored to specific project conditions.

This review was not designed to be an exhaustive treatise on evaluation design in general. As such, there is no discussion on design issues such as experimental design vs. quasi- or non-experimental design, sampling design and frameworks, or on the pros and cons of various methods and tools for evaluation. These issues have been discussed in the Approach Paper for GEF impact evaluation.¹

Modes of Impact Evaluation

Before we move much deeper into the analysis found in this report, we feel it is important to clearly define some key terms related to impact evaluation. The Organization for

¹ Todd, David and Josh Brann. GEF Evaluation Office, February 6th, 2006. *GEF Impact Evaluations, Initiation and Pilot Phase - FY06: Final Draft Approach Paper*

Economic Cooperation and Development's (OECD) Development Assistance Committee (DAC)² defines *impacts* as “Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended” and the Operations Evaluation Department of the World Bank defines *impact evaluation*³ as the systematic identification of these effects. More specifically, it refers to impact evaluation as:

- The process of determining the worth or significance of an activity, policy or program⁴; and
- An assessment, as systematic and objective as possible, of a planned, on-going, or completed development intervention.⁵

While we agree that these definitions go a long way to describe what needs to be accomplished in order to measure the impact of GEF interventions, we believe an even more precise and specific definition is in order for this study. For the purposes of the work described in this report, therefore, we use the following definition of impact evaluation:

Impact evaluations assess program effectiveness in terms of end or ultimate results, including those that are intended and unintended. An impact evaluation should analyze the extent to which intermediate results – or outcomes – occur, but this is not sufficient to gauge impact. It must also assess the degree to which the ultimate desired result has been achieved. Impact evaluation is possible only when a strong likelihood exists that the intervention contributes to the long-term changes in the desired impacts, as through a series of “if-then” relationships.

In addition to an evaluation of impacts as described above, we believe that GEF evaluations must focus on five additional critical components, including:

- **Sociopolitical sustainability** – The extent to which project activities lead to long term improvements in the social and political situation where the project is found and where such changes are essential to ensure improved environmental management.
- **Programmatic sustainability** – The extent to which the actions that are taken during the life of the project continue after the formal project ends.
- **Institutional sustainability** – The extent to which necessary institutional structures are in place and secure for the long term as a result of the project.

² www.oecd.org/dac

³ World Bank. *Monitoring and Evaluation: Some Tools, Methods & Approaches*.

<http://Inweb18.worldbank.org/oed>

⁴ www.worldbank.org/ieg/eed/what_is_me.html

⁵ www.worldbank.org/oed/ipdet/modules/M_01-na.pdf.

- **Financial sustainability** – The extent to which post-project activities can sustain themselves financially or mechanisms are in place to provide a constant flow of external financial resources.
- **Replication** – The extent to which successful implementation of actions in one project can be repeated in other project sites.

In its *Final Draft Approach Paper*, the GEF EO describes a spectrum of possible approaches to impact evaluation, based on current practices. These include:

The Counterfactual-Based Approach

According to this approach, attribution requires an explicit “counterfactual,” which establishes what would have happened if the intervention had not been made. The impact of a project can then be assessed as the difference between the observed outcome and the counterfactual. While believed by some to be the most objective approach, it tends to be time, money and skills intensive, and very difficult to do well.

Narrative/Historical Evaluation (Including “Shoestring” approaches)

This approach involves a full description of the external intervention or interventions, coupled with an assessment of changes in development or environmental status which appear to have been associated with it. The intent is to show the extent to which the observed changes can be attributed to the interventions being evaluated. This approach works under the assumption that it is more appropriate to use impact evaluations to identify an intervention’s contribution to observed changes, rather than to attempt to establish causality.

Theory Based Approaches

Theory based evaluation follows a logical sequence of cause and effect linkages in which the evaluation explores the extent to which events followed the anticipated sequence and the intervention achieved the desired objective. The evaluation assesses the validity of the theory of change adopted (implicitly or explicitly) by the intervention, as well as the interaction between the specific local circumstances and the general principles expected to generate the desired change. The approach enables detailed examination of the nature of linkages between a complex set of causes and a set of effects. The theoretical model can clarify the assumptions upon which the intervention was based and also incorporate opposing theories, which may be applied to the same situation.

Based on its initial analysis, the GEF EO has decided to take a theory-based approach to developing the evaluation framework. By doing so, the EO wishes to analyze expected and actual results along a theory of change sequence. At the project level, this means attempting to answer these key questions outlined in the terms of reference for this work:

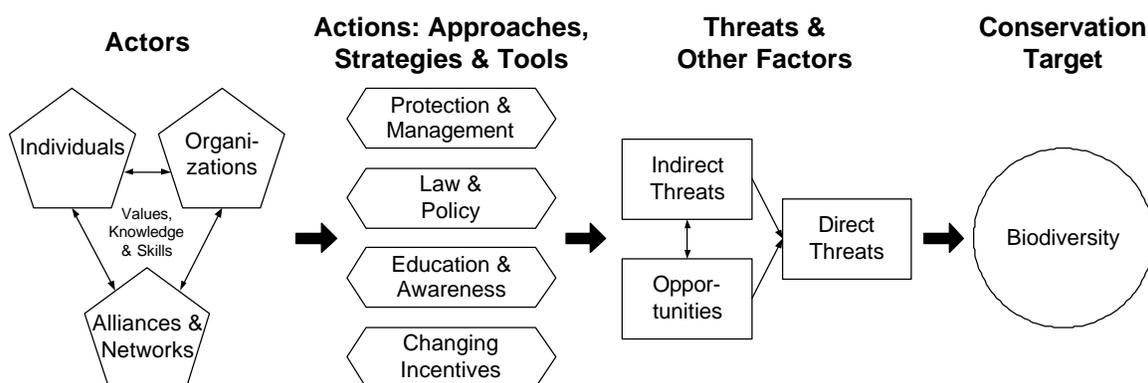
- What was the intended series of cause and effect linkages (mechanisms), which were expected to generate impacts?
- What were the key features of the project context, which interacted with these linkages to determine results achieved (outputs, outcomes, and impacts)
- How did the project respond to its specific context to generate results?

- How do impacts at the project level relate to global environmental status and the overall objectives of the GEF?
- Are additional results from the project anticipated in the future, and if so, to what extent?
- Are project results likely to be sustained?
- What does the project tell us about the underlying theory of change on which the intervention was based?
- What does the project tell us about the interaction between the intended change mechanisms and project context, which could form the basis of lessons for future interventions?
- What lessons does the evaluation present concerning possible improvements to impact evaluation methodology?

Evaluation State-of-the-Art

Increasingly, the biodiversity community is coming to consensus on the basic structure and underlying logic of conservation projects (Figure 1). There is an emerging and general consensus that conservation projects generally involve four components: a conservation target, threats and opportunities, actions, and actors. A conservation target is the specific aspect of biodiversity a project is designed to influence. This can include, for example, species, habitat or ecosystems. Threats are those human induced pressures that adversely affect the conservation target. Opportunities are factors that favorably influence the conservation target. Actions are the interventions that are designed to mitigate a threat or support an opportunity. Actors include the people and institutions that design and implement actions. In sum, managers usually design project actions to reduce one or more threats or support opportunities with the intention of having some positive impact on a given conservation target.

Figure 1: Basic Structure of a Conservation Project



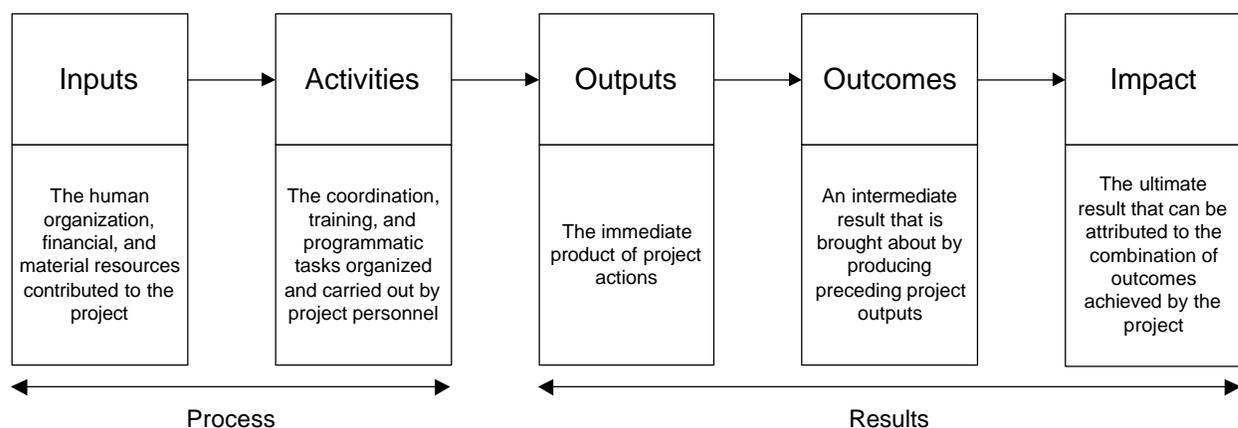
In generic terms, conservation projects are designed to influence a series of cause-and-effect relationships that affect conservation targets. This generic theory of change

illustrates what project managers plan to do (the *process*) and what they expect to achieve (the *results*). In the first segment of the model, project managers invest resources (*inputs*) – including budget, staff time, and materials – to support their *activities*. Activities include the specific programmatic actions described in the above diagram and all other related activities such as administrative and financial support, management training, and logistic support.

According to this construct, this process leads to three levels of results, represented in Figure 2. These concepts and term are widely accepted by many disciplines:

- **Outputs** – are the immediate products of a given action or intervention – they are the quantity of goods and services produced and the efficiency of production. Examples of outputs include the number of people reached by an environmental education campaign and the percent of a core area boundary that has been demarcated by park guards.
- **Outcomes** – are the results that are reflected in the threat or opportunity a project is designed to address. They are the product of addressing the project objectives. They represent, for example, % of hunters no longer hunting illegally, presence of exotic species X in a given area, and rate of tourist vehicles that go off-road inside a park.
- **Impact** – is the ultimate result of a project that can be measured at the level of the conservation or global target. Assessment of impact is the measurement of the effects on stated project goals. Examples include population structure of mountain gorillas in Park X and distribution of plant species Y in a given area.

Figure 2: Project Process and Results Components



Evaluation is useful to project and program managers to determine the extent to which their conservation actions are working and to make necessary adjustments if they are not.

Under ideal conditions, evaluators would want to collect data on inputs, activities, outputs, outcomes, and impacts. Related specifically to our model of a typical conservation project, this means they would want to collect data related to actors, actions, threats/opportunities, and conservation targets.

But the simplified model we present above belies the inherently complex nature of biodiversity conservation actions and the contexts within which they take place. In conservation projects – as in many other types of projects – there are two main types of complexity. Detail complexity refers to the large number of variables in the system. Dynamic complexity refers to the unpredictable ways in which variables interact with one another. Since conservation involves combining both natural ecosystems and human societies, we are dealing with systems that are inherently extremely complex both in detail and in dynamic. One set of challenges to establish a meaningful evaluation approach is to understand and manage these types of complexity.

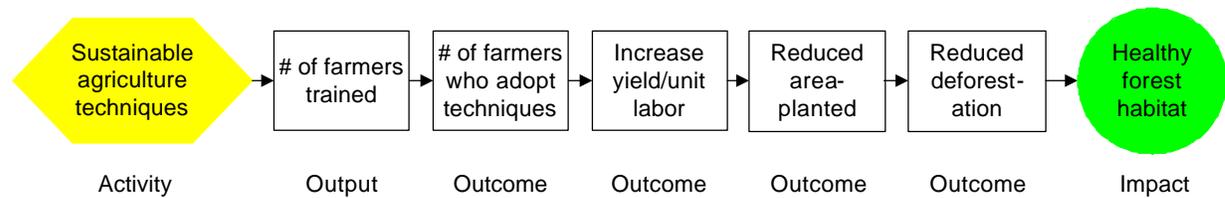
It is impractical to believe that we will ever be able to completely eliminate detail or dynamic complexity – that we will understand all variables and the interactions among them – that influence conservation results. But it is an underlying assumption of Theory of Change models that we can significantly reduce the effects of these types of complexity by understanding them better. The greater our success in understanding this complexity, the higher the probability that we will be able to design, implement, and monitor conservation projects that are more likely to succeed.

In addition to the challenges presented by detail and dynamic complexity, there are real and very practical limits to what data are feasible to collect and use for evaluation. The magnitude of these limits differ from project to project, but invariably, time and budget constraints mean that project managers must be selective when deciding what to measure for evaluation purposes.

In order for us to propose a plausible approach to evaluating GEF projects, we need to be able to guide evaluators to identify and select the best data and information to collect under different project conditions. To do this, we need to assist evaluators to understand the assumed causal mechanisms that lead from action to results, and thus, be able to understand how relevant indicators change depending on outside forces, including a project's own actions.

Understanding the assumed causal mechanisms that lead from a given conservation action to specific results is the first step in a Theory of Change approach to evaluation. Figures 1 and 2 above have provided generic representations of assumed causality. Figure 3 provides a concrete example of a conservation project that employs a theory of change related to sustainable agriculture as a tool to reduce deforestation rates and ultimately lead to healthy forest habitat. In this example, multiple outcomes are generated that are assumed to lead to biodiversity conservation impact.

Figure 3: Theory of Change for Sustainable Agriculture



Once an evaluator understands the theory of change between a given action and a given impact, he/she is in a better position to identify and select appropriate indicators to assess change. If we were to establish priorities for identifying the most important categories of evaluation indicators, the highest priorities for the above models would probably go from right to left – that is, targets/impacts are the most important, followed by threats and opportunities/outcomes, followed by actions/outputs and activities, followed by actors/inputs. It thus makes sense that much attention is paid to the collection of data related to targets/impacts.

But there are some inherent difficulties in relying solely on impact indicators to measure project success. First, as one moves further down the results chain from action to impact, the causal relationship becomes weaker and weaker, as many other factors come into play and influence the target. Thus, although you may be able to measure a target, your ability to attribute some measured change in the target to your intervention becomes increasingly difficult.

Establishing the assumed causal link at this point is critical, and one way an evaluator can increase his/her ability to decipher this link is to measure an expected intermediate result – outcome indicators related to threats and opportunities. If an evaluator can show that a given project implemented activities that produced an output that led to a change in threat status, and in turn the evaluation uncovered some expected change in the conservation target, then the evaluator can be more confident that the intervention contributed to this change. Directly measuring activities, inputs, and outputs is a far less challenging endeavor as project managers can easily record these as they occur. They require, therefore, significantly less attention and resources than the measurement of outputs, outcomes and impact. In addition, by nature, measuring indicators for inputs and activities is far more straightforward than measuring those for outcomes and impacts.

From a purely programmatic point of view, the most important yet difficult measurements to make are of outcomes and, to an even greater extent, impacts. But from an administrative or management perspective, it may be equally important to understand returns on investments. Most project managers want to know how they can spend the least amount of their budget to achieve the greatest amount of results. In order to do this, inputs data – like cost and staff time – are extremely important to collect.

In sum, although the highest priority in an evaluation approach may be identification and gathering of output, outcome and impact data, data related to inputs and activities should be collected as needed in order to complete the theory of change and to clarify attribution.

Types of Theories of Change Tools

The name “Theory of Change” is a generic or umbrella title given to a group of tools that are used to make explicit underlying assumptions that link actions to impacts. The Aspen Institute’s Theory of Change website⁶ provides a good introduction to this concept:

“Theories of change have been largely used as a tool for evaluation. It is probably impossible to pinpoint the “first” use of the term “Theory of Change,” but a hint at its origins can be found in the evaluation community among the work of notable methodologists, such as Huey Chen, Peter Rossi, Michael Quinn Patton, and Carol Weiss. These methodologists, along with a host of others, have been thinking about how to apply program theories to evaluation for at least 20 years.

Carol Weiss popularized the term “Theory of Change” as a way to describe the set of assumptions that explain both the mini-steps that lead to the long term goal of interest and the connections between program activities and outcomes that occur at each step of the way.

As we define it, a Theory of Change defines all building blocks required to bring about a given long-term goal. This set of connected building blocks – referred to as [outcomes](#), results, accomplishments, or preconditions – is depicted on a map known as a pathway of change/[change framework](#), which is a graphic representation of the change process.

Built around the pathway of change, a Theory of Change describes the types of [interventions](#) (a single program or a comprehensive community initiative) that bring about the outcomes depicted in the pathway of a change map. Each outcome in the pathway of change is tied to an intervention, revealing the often complex web of activity that is required to bring about change.

A Theory of Change would not be complete without an articulation of the [assumptions](#) that stakeholders use to explain the change process represented by the change framework. Assumptions explain both the connections between early, intermediate and long term outcomes and the expectations about how and why proposed interventions will bring them about. Often, assumptions are supported by research, strengthening the case to be made about the plausibility of theory and the likelihood that stated goals will be accomplished.”

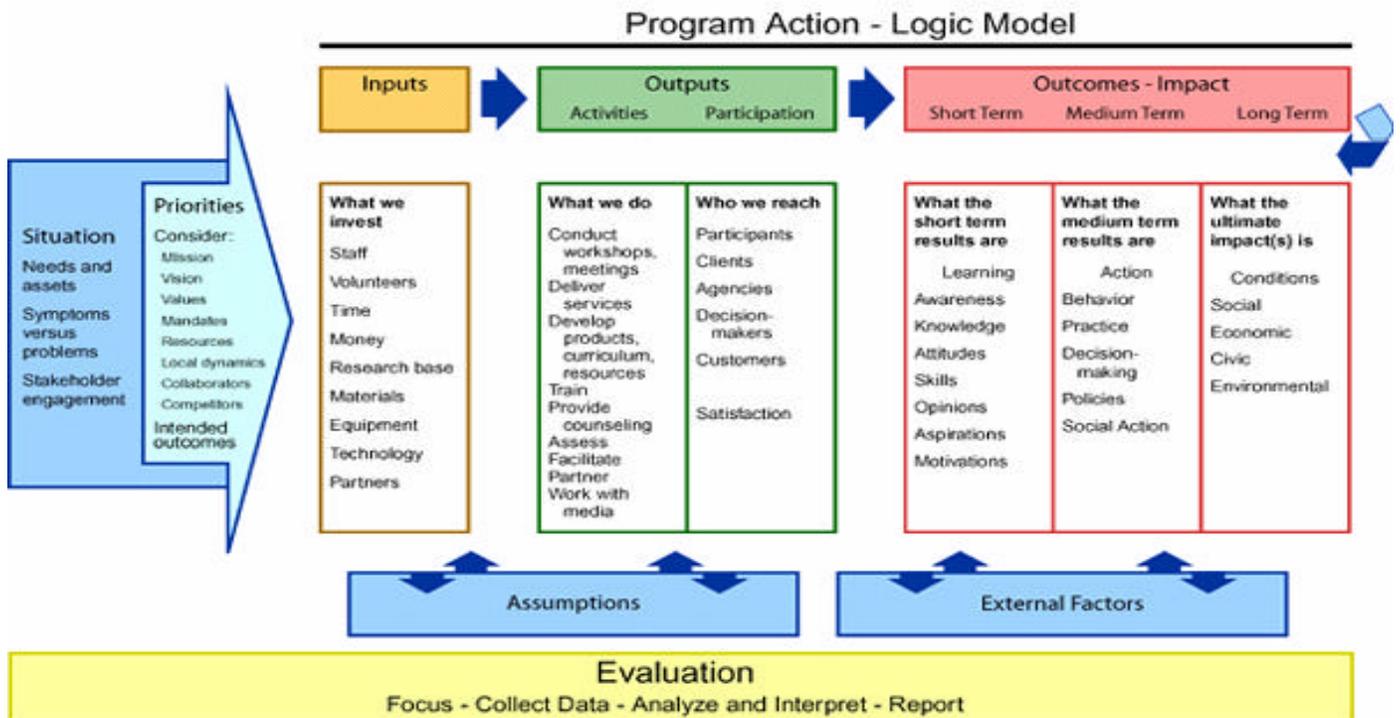
The most commonly used tools associated with a Theory of Change approach are logic models and results chains – or some slight variation of these. Because they are so widely used and have been extensively tested in many fields, these are the tools that we have relied on for this work.

⁶ <http://www.theoryofchange.org/index.html>

Logic Models

Logic models (Figure 4) are a general and yet systematic and visual way to present the perceived relationships among the resources used to operate the program, the activities undertaken, and the intended changes or results. Logic models are generally fairly easy to do, but their level of precision is low. Furthermore, logic models tend not to show specific linkages among specific actions, threats, and targets.

Figure 4: Generic Logic Model



Results Chains

In general, results chains are a tool that clarifies assumptions about how conservation actions contribute to reducing threats and achieving the conservation of biodiversity targets (see Figures 5a and 5b). In particular, a results chain:

- Is a diagram of a series of “if...then” statements (“causal”)
- Defines how we think a project strategy or activity is going to contribute to reaching a goal
- Focuses on the achievement of results – not the execution of activities
- Is composed of assumptions that can be tested

There is a series of criteria that must be fulfilled in order to fully develop a results chain. These include:

- Results oriented

- Simple
- Connected in a “causal” manner
- Demonstrates changes
- Relatively complete

Figure 5a: Generic Results Chain

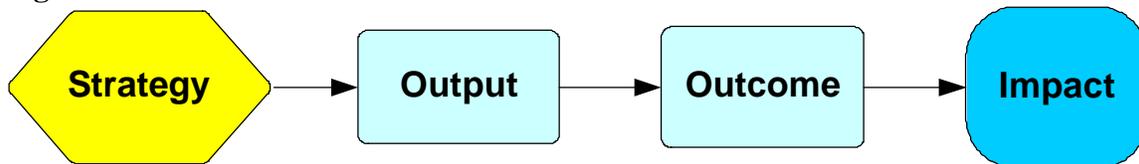
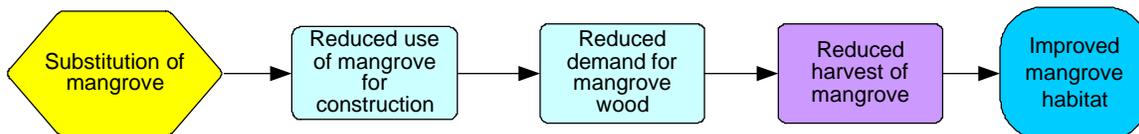


Figure 5a: Example Results Chain



In general, results chains are much more specific and precise than logic models in that they show the precise relationships between expected results, i.e., along the chain. Ideally in all projects, therefore, evaluators’ first choice should be to try to develop results chains as the basis for the evaluation. If, however, there simply is not enough information to create results chains, then the evaluator should go to the more generic and less precise logic model.

Methods

The main focus of this work was to develop an evaluation approach suitable to the needs of the GEF Evaluation Office (EO). To do this, we reviewed project documents from a sample of GEF-funded projects identified by the EO team. We wished to create theories of change that were evident in project documents and that might represent prevailing priority programmatic approaches identified by the GEF Secretariat. There are two reasons for this: First, we wanted to determine if we could actually evaluate each specific project using a Theory of Change approach. To do this, we would have to reconstruct the logic used to design the project and assess the extent to which data were collected to measure progress along the relevant results chain. Second, in order to assess whether evaluators could aggregate GEF projects in meaningful ways that would lead to program-level (vs. project-level) evaluation, we wished to determine if there were common theories of change *among* projects, or some other meaningful way of organizing similar projects. In addition, we expected that this aggregation of theories of change among

projects would enable us to determine the extent to which it is feasible to expect that GEF can “roll-up” results and capture some common lessons-learned across similar projects.

We anticipated that few projects would have clearly delineated theories of change – logic models or results chains – articulated in project documents and reports. Instead, we expected that we would have to retrospectively recreate chains based on what we can interpret from project documents. In fact, that was the case: we interpolated most of the expected causal linkages from background information, analysis of threats, and activity descriptions contained in project documents.

Sample

The EO decided early on that it wished to focus this feasibility study and a subsequent pilot study on the biodiversity focal area. It did this because the Climate Change portfolio has figured heavily in recent World Bank and GEF impact studies and the International Water portfolio is likely to have a long lead time, before an impact evaluation approach can be applied.

The EO selected a sample of thirty projects from the biodiversity focal area: twenty pre-GEF-3 projects for which Terminal Evaluations were completed and ten GEF-3 projects. To the extent possible, this sample was evenly distributed across IAs and regions. In the end, FOS analyzed ten pre-GEF-3 projects and ten GEF-3 projects. For pre-GEF-3 projects, we reviewed primarily Project Documents and Terminal Evaluations. For GEF-3 projects, Terminal Evaluations had not been done. (See Attachment 1 for a list of the overall sample and the projects that were actually reviewed for this work.)

Determining Utility of Approach to Evaluate

In order to determine the extent to which a Theory of Change approach could work as an effective mechanism to evaluate individual GEF projects, we reviewed twenty projects from the original sample and attempted to develop logic models and results chains for key strategies identified in project documents. We reviewed all available documentation on the GEF website including Terminal Evaluations for all Pre-GEF-3 projects. We combed these documents to extract any piece of information that would help us construct either logic models or results chains. In most cases, we had to interpret results and assumptions because they were not explicitly articulated in the project documentation. We attempted to identify all project strategies for each of the projects we reviewed and we developed logic models and results chains for a subset of these strategies for each project.

Determining Utility of Approach to Aggregate Results

To gauge the extent to which a Theory of Change approach lends itself to the aggregation of results in a diverse portfolio such as the biodiversity focal area, we used the basic structure of a conservation project in Figure 1 – including actors, actions, threats/opportunities, and targets – as a way of organizing our analysis. We also used the IUCN/Conservation Measures Partnership (CMP) Classification of Conservation Actions and Threats⁷ to order the respective components of each project.

⁷ Available at: http://conservationmeasures.org/CMP/IUCN/Site_Page.cfm

Using an extensive Excel spread sheet, we first entered all relevant data for each project in order to compare them on the individual components described in Figure 1. On one axis, we included the projects, and on the other axis, we included the component categories of the IUCN/CMP classifications. In this way, for example, we could determine how many projects addressed a particular threat, or implemented a particular strategy.

We then looked across rows in the spreadsheet to determine how many times – out of a possible twenty projects – a particular strategy, threat, or target was mentioned. But we took this analysis one step further by looking across rows and projects to determine the extent to which the same combination of actions, threats, and targets could be found in various projects.

Findings

Based on our review of the twenty projects and countless reports, interviews with key individuals, and discussions with EO Staff, we support the preliminary conclusion of the EO that a Theory of Change approach to impact evaluation provides a sound basis from which to evaluate GEF projects. It also makes sense given the advances that have been made in GEF project design, management, and reporting in recent years. In particular, with the inclusion of logframes beginning with GEF-3 projects, most of the more recent projects are set up to be evaluated using a Theory of Change approach. In the following sections, we present a summary of some key observations of the twenty projects we reviewed. Our observations fall into two rough classifications: pre-GEF-3 projects, and GEF-3 projects. For the latter, the results apply to GEF-3 and subsequent phases so we refer to these as “GEF-3+” projects.

Analysis of Pre-GEF-3 Projects

Of these older GEF projects (which include terminal evaluations), about half include enough desired project results and outcomes (usually stated as objectives, outputs and activities) to be able to construct at least simple theories of change. There are, however, inconsistent use and definition of terms and concepts in the project documents we reviewed.

Many of the projects have gaps in their logic. For example, one of the theories of change of the Lebanon protected areas project (Strengthening of National Capacity and Grassroots In-Situ Conservation for Sustainable Biodiversity Protection) is that an environmental education project will sensitize local people to the importance of conservation and increase dialogue about resource conflicts, which will reduce threats to the protected areas. The project does not seem to identify or address the underlying causes of unsustainable resource use activities or otherwise what is expected to happen between opening of dialogue and reduced threats (e.g., increased trust or improved understanding).⁸

With some complex strategies, such as conservation trust funds, many of the project proposals do not seem to recognize how much time it takes to implement the strategy, or what institutional capacity is needed as a prerequisite, in order for the strategy to be successful. For example, many of the older GEF projects include a conservation trust fund as a mechanism to increase the financial sustainability of the project, but only a few projects (such as the Belize Conservation and Sustainable Use of the Barrier Reef Complex project) seem to understand what the process of establishing a trust fund entails and what some of the outcomes are.

⁸ What this analysis tells us about the proposed impact evaluation methodology is that projects should be evaluated on the stated or implied logic, but this does not preclude the evaluator from suggesting that the logic was faulty and perhaps other things – such as increased trust and improved understanding as per this example – were needed but not addressed by the project.

Proposals for projects being established in countries that have little experience in biodiversity conservation (e.g., Conservation of Biodiversity and Sustainable Development in the Federal Islamic Republic of Comoros and Lao PDR's Wildlife and Protected Areas Conservation project) tend to be very general and have weak programmatic logic. They also tend to be overly ambitious (e.g., Lao) and not accomplish many of their objectives – probably in large part due to weak program design. Lessons from other countries that have already gone through these early stages of developing the capacity to manage biodiversity projects do not seem to have been transferred to these countries.

Some of the projects examined seem to try to do everything at the same time. Developing theories of change for projects like these would be quite difficult, because they do not go into enough detail about each of the individual activities and its expected results. Projects that focus on fewer strategies and develop them more fully lend themselves better to the Theory of Change approach.

Some of these projects are designed at a macro scale to improve the management and long-term financing of several protected areas (PAs) – 17 in the case of the Mexico Protected Areas Program project. Developing theories of change for this project would require understanding what the targets, threats, and management activities are in those specific PAs – a level of detail not included in the project documents examined, and potentially difficult for an outside evaluator to obtain. The project was considered “extraordinarily successful,” based on some indicators measured in all PAs (e.g., occurrence of endemic or endangered species and reduction in habitat conversion) and the fact that all PAs met 80-100% of the norms for technical and financial implementation (process indicators). But the question remains: To what extent can GEF claim that it had any influence on conservation impacts? What evidence exists that links interventions to impacts?

In several GEF documents – especially UNDP projects – indicators are stated as objectives (because they include results they want to achieve in the indicator itself). This type of definition of objectives is extremely difficult to evaluate.

Reported results in project documents often vary from outputs (e.g., management plans developed) to threat reduction goals (e.g., agricultural encroachment reduced by 50%) and the relationship between different types of results is often not clearly defined. In addition, project documents seldom provide justification given for the designated threshold levels for objectives and goals. This makes it easy to inflate expected results without clearly articulating how feasible or accurate they are. See Box 1 for an example of this in Guatemala.

One observation that applies to several projects reviewed is that GEF project managers and evaluators appear to claim results for which the project made a minor contribution or was but one of many actors contributing to the results. Again, see Box 1 for an example of this in Guatemala.

Similarly, reported project results often cannot be logically or programmatically justified in many projects. For example, in the China Natural Reserves Management project, participating reserves reported increases in sightings of pandas, golden monkeys, and golden takins, increases in vegetative cover, and decreases in key threats (e.g., logging, hunting, and fires) – and attributed these directly to the project. But there is no evidence that these conclusions are valid, that they are causally linked or associated, or that they can be realistically attained within the timeframe of the project.

Box 1. An Informed Outsider's Perspective on the Guatemala RECOSMO Evaluation

The evaluation of the Guatemalan RECOSMO project illustrates some of the concerns that emerged through the present study to propose an evaluation approach for the GEF. Here, we outline some of those, although in doing so, we do not mean to single out the RECOSMO project but rather to use it to illustrate more concretely some of this study's findings.

Lack of Explicit Assumptions about Theories of Change

The evaluation of the Guatemalan project (Integrated Biodiversity Protection in the Sarstun-Motagua Region) noted that the goals of the subsequent project (RECOSMO II) were very similar to the original RECOSMO goals and the expected and obtained results were too vague. This is probably in large part because they do not explain their theories of change or how far along that theory of change they moved after RECOSMO I.

Lack of Clarity about Setting Thresholds

In addition, the project was not clear about how it set its threshold levels for its goals and objectives, leading to a tendency to overestimate expected results. For example, the goal was to involve 30% of the population in activities compatible with environmental conservation. The project documentation did not discuss how the value of 30% was decided. The consultant's detailed knowledge of this particular site suggests that this is not even a remotely feasible goal.

Limited Feedback and Use of Evaluation Results

The evaluation of the Guatemala project noted that the project did not have adequate mechanisms in place for using M&E results or providing feedback to national implementing organizations. The evaluation also included the observation that external evaluations were not frequent enough to provide feedback necessary to correct deficiencies in the project. The evaluators concluded that an internal M&E system would have been more effective.

Misleading Attribution of Project Results

The consultants are aware that GEF funding for the RECOSMO project was in the form of relatively small contributions to many different and divergent activities. Nevertheless, the final evaluation gives credit to GEF for some results actually primarily supported by other donors. One example of this is the Sierra de las Minas water fund in which GEF is one of about nine funding sources. In the Terminal Evaluation, there is no acknowledgement of contributions of other projects or donors before or during RECOSMO. Similarly, the evaluation rates the Sierra de las Minas and Cerro San Gil protected areas as the best managed PAs, but it does not even mention that national NGO Defensores de la Naturaleza has been managing the Sierra de las Minas Biosphere Reserve for over fifteen years and national NGO FUNDAECO has been working in Cerro San Gil for over ten years – again, in both cases, with the lion's share of support coming from other projects and donors. There is simply no acknowledgement of the level of effort, time, or resources needed to accomplish results claimed by GEF, how GEF's contribution falls within a longer-term context, and that GEF support builds on extensive management capacity in and support to existing PAs.

Analysis of GEF-3+ Projects

Almost all of the GEF-3+ project documents have identified the short-term and intermediate outcomes that they believe are needed for their project activities to lead to threat reduction and conservation of ecosystems and species. Of the ten projects reviewed, nine of them included enough outcomes to be able to develop sufficiently detailed theories of change. Only the Brazil project – National Biodiversity Mainstreaming and Institutional Consolidation Project – lacked the necessary information.

Most of the required details to construct theories of change can be found in project logframes. However, there is some inconsistency in the formats of the logframes that are used across IAs. This can make comparisons across projects difficult. For example, the logframe for the India project – Mainstreaming Conservation and Sustainable Use of Medicinal Plant Diversity in Three Indian States – includes project strategy (goal/objectives), indicators, baseline level of indicators, target level of indicators, sources of verification and risks and assumptions. The Honduras logframe (for the Consolidation of Ecosystem Management and Biodiversity Conservation of the Bay Islands project) does not include the baseline or target levels for indicators. It includes narrative summary (goals/objectives), verifiable indicators, means of measurement, and assumptions. The Tanzanian Marine and Coastal Environmental Management Project uses a results framework that includes – at the top of the table – the project development objective/intermediate results, outcome/results indicators and use of outcome/results indicators. The left-hand column of the logframes and results frameworks sometime include the project goal and objectives and sometimes include results or objectives and outcomes.

While much of the raw materials can be found in the logframes, most actual project objectives do not meet SMART (specific, measurable, achievable, realistic, and time-bound) criteria – in particular, most are not specific or measurable.

Within the logframes, there is little consistency in the application of terms. For example, while the India project includes specific indicators such as “Hectares of forest actively managed for sustainable use of MAPs and maintenance of MAP diversity,” the Honduras project includes the following as an indicator: “Improved environmental conditions and recruitment in coral reefs, fish and crustacea stocks, leading to improved ecosystem functions in the Bay Islands”, which meets criteria most closely resembling a goal.

Projects operating on a very broad scale tend to have weaker project logic. For example, the Brazilian project on mainstreaming national biodiversity includes the following objectives: “to mainstream the conservation and sustainable use of biodiversity into select economic sectors at federal and state government levels” and “to mainstream the conservation and sustainable use of biodiversity in the private sector.” The project seems to be trying to address almost all of the underlying causes of biodiversity loss in Brazil, rather than defining and clearly focusing on some of the most important underlying

threats, such as the agricultural development policies and financing for agricultural expansion that represent a major underlying cause of habitat conversion and degradation.

Some of the projects' theories of change are defined in the project documents but include some large leaps of faith (big jumps from one outcome to another) -- for example, many projects include the following theory: small-scale, sustainable economic activities lead to poverty reduction, which leads to reduction in destructive resource use activities, which leads to greater conservation of ecosystems and species. Another example is the Tanzania project's belief that educational campaigns lead to the elimination of destructive practices.

While it is often possible to identify the intended project outcomes, the project documents do not always "connect the dots" between these outcomes and threat reduction and biodiversity target conservation. Biodiversity targets are often described in one part of the document, threats to biodiversity in another part (and often not linked to specific conservation targets), and strategies in yet another part and usually not clearly linked to reduction of specific threats. The logframes generally focus on the left-hand side of results chains – i.e., strategies/activities, outputs, and immediate outcomes. Some of the better logframes (e.g., India, Belarus) also include the conservation targets.

In the conservation community, it is common for conservation projects to talk in general terms about developing models – and the replication of these models elsewhere – without thinking about the mechanisms needed for that replication to happen. Some of the GEF-3+ projects reviewed, however, included specific outcomes needed to ensure replication.

Applying a Theory of Change Approach to the Sample

Perhaps because GEF has relatively recently shifted towards logframes and associated tools, it is more likely that GEF-3+ projects have sufficient logical "raw materials" to be able to conduct a Theory of Change evaluation. Equally important, even for those pre-GEF-3 projects that did not enjoy the benefit of a logframe, it appears that there are at least some basic ingredients present to construct a more simplified Theory of Change. There is, however, one note of caution based on our analysis: while there is wider adoption of logframes across the GEF portfolio in more recent years, the actual structure of the logframes and definition of key terms vary across IAs.

While we found basic elements of theories of change in just about all of the projects we reviewed, the possibility for evaluators to retroactively develop theories of change will vary by project. Their possibilities for apply specific tools related to theories of change (e.g., logic models and results chains) will also vary. When conducting evaluations, therefore, evaluators will need to spend varying amounts of time and effort organizing information and structuring expected results in order to adapt them to a Theory of Change approach.

We took the sample of projects reviewed for this study and attempted to apply a Theory of Change approach, using logic models and results chains to represent major strategies. The purpose of this exercise was to determine if we could use theories of change to set up

the evaluation, and if so, what level of effort might be required to do so. We hoped to be able to apply the much more specific and explicit results chain approach to all projects, but given that some projects did not define results clearly or did not express underlying assumptions, we were not entirely successful in doing this.

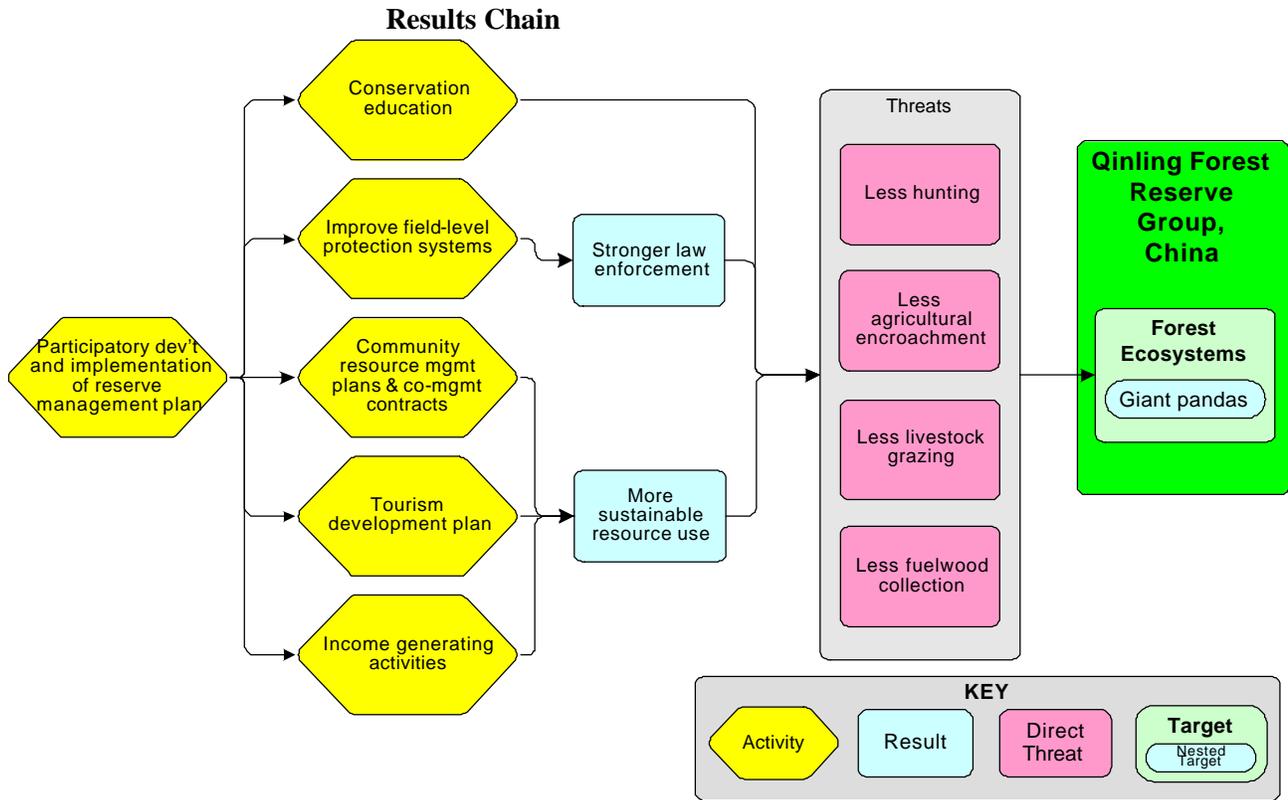
For each of the three cases we present below, we chose a specific project from our sample and applied the logic model and the results chain approaches to describing theories of change. The sample was purposeful – we selected one case that is a pre-GEF project strategy for which there is little explicit articulation of results; another case that is a pre-GEF project strategy that lends itself to both logic models and results chains; and a final case that is a GEF-3+ project strategy that has ample information to set up both logic models and results chains. The purpose of this exercise was to assess the conditions under which each tool could be used in a Theory of Change evaluation approach

Case 1: Protected Areas Management Strategy – China Nature Reserves Management/pre-GEF-3

In this example of a pre-GEF-3 project strategy, the logic model does an adequate job of describing – in fairly general terms – the relationship between inputs, outputs, outcomes and impacts. However, analyzing the results chain, which should provide more detail about necessary intermediate results that lead to a desired outcome, there is very little information provided by project documentation that can be used to construct a sound theory of change. In this case, most of the factors in the chain are either actions or threats, with little understanding of the conceptual or programmatic linkages that connect them. In sum, for this particular strategy, the logic model works fairly well, but the results chain approach does not.

Logic Model

Inputs	Outputs	Results		
		Short-term Outcomes	Intermediate Outcomes	Impacts
<ul style="list-style-type: none"> • Financial resources • Human resources • Equipment 	<ul style="list-style-type: none"> • Participatory preparation and implementation of reserve and land use management plans • Improving field level protection systems • Development of community resource management plans and co-management contracts between communities and reserves • Development of income generating activities for communities • Preparation of tourism development plans • Conservation education 	<ul style="list-style-type: none"> • stronger law enforcement • more sustainable resource use 	<ul style="list-style-type: none"> • Less agricultural encroachment • Less livestock grazing • Less hunting • Less fuelwood collection 	<ul style="list-style-type: none"> • Greater conservation of forest ecosystems in reserves and in biological corridor linking reserves • Greater conservation of giant panda habitat and giant panda population



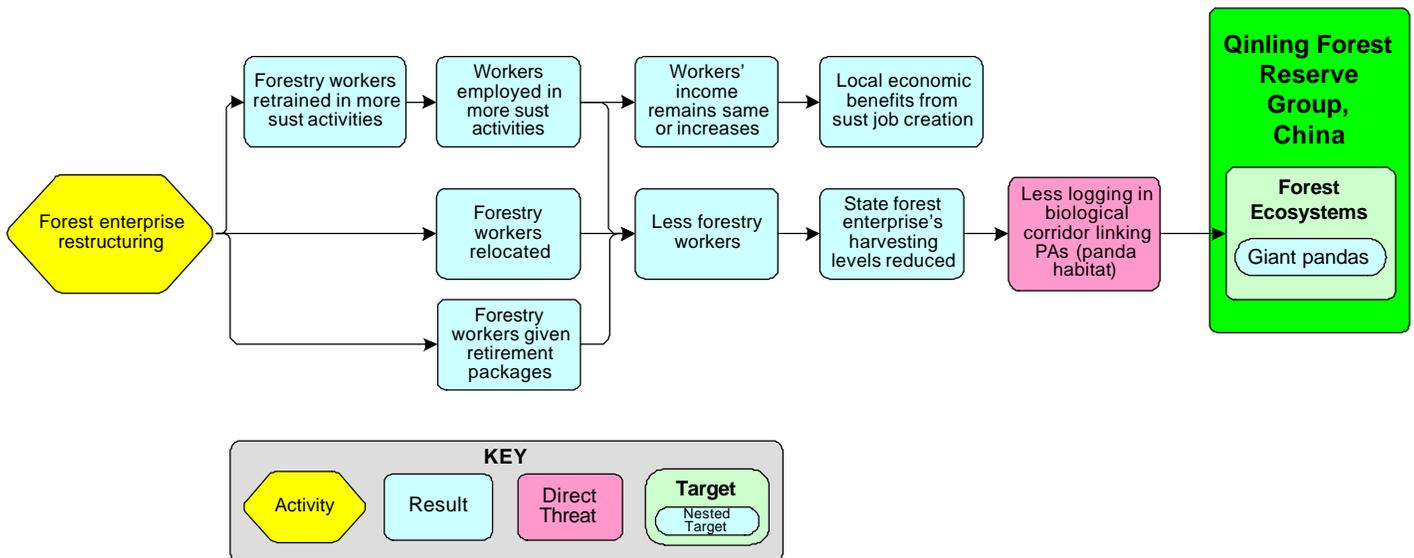
Case 2: Forest Enterprise Restructuring Strategy – China Nature Reserves Management /pre-GEF-3

In this case, also a strategy from the China pre-GEF-3 project, both the logic model and the results chain do an adequate job of representing the underlying logic of the project. As expected, the logic model is general, but the results chain provides a wealth of information – particularly in the intermediate results area between action and threats.

Logic Model

Inputs	Outputs	Results		
		Short-term Outcomes	Intermediate Outcomes	Impacts
<ul style="list-style-type: none"> • Financial resources • Human resources • Equipment 	Forest enterprise restructuring	<ul style="list-style-type: none"> • Forest workers given retirement packages, relocated or retrained and employed in more sustainable activities • State forest enterprise’s harvesting levels reduced 	Less logging by state-owned forest farms in buffer zone of reserve (biological corridor for pandas)	<ul style="list-style-type: none"> • Greater conservation of forest ecosystems in reserves and in biological corridor linking reserves • Greater conservation of giant panda habitat and giant panda population

Results Chain



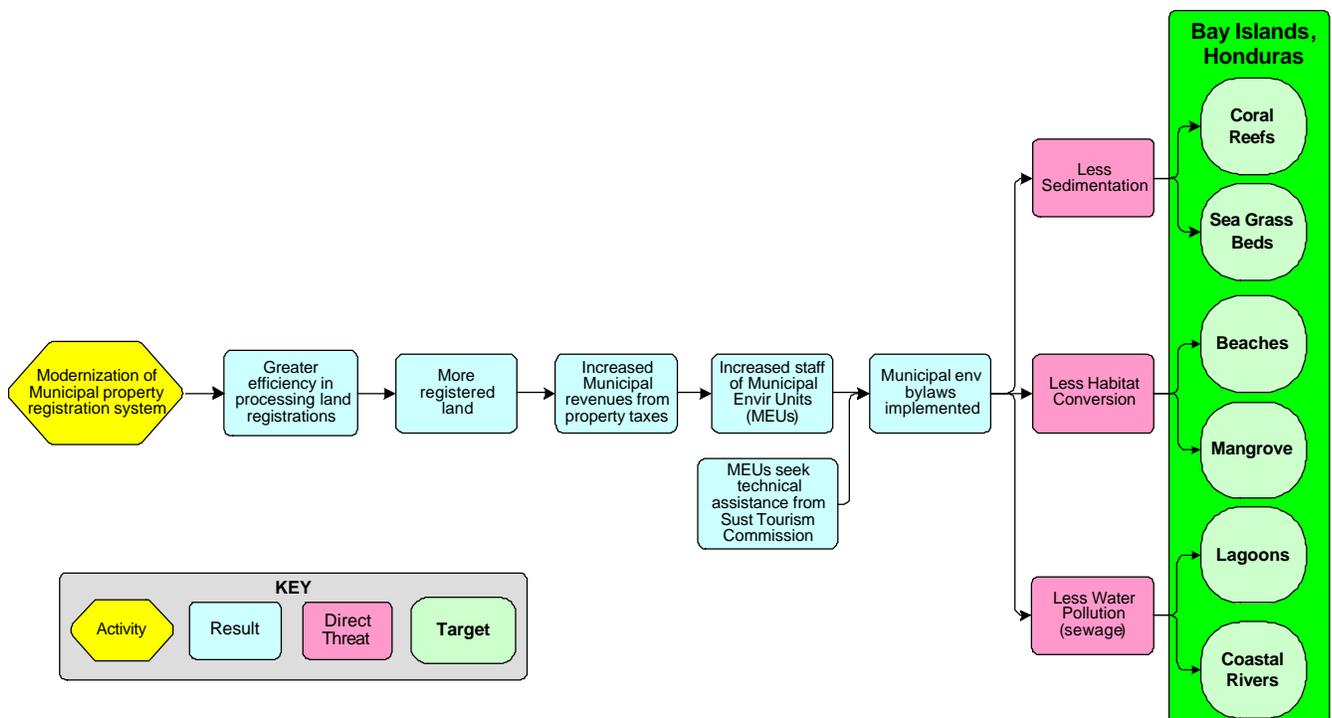
Case 3: Strengthening the Environmental Management Capacity of Municipalities – Honduras Consolidation of Ecosystem Management and Biodiversity Conservation of the Bay Islands/GEF-3+

Not surprisingly, in this GEF-3+ case, there is more than ample information available from project documents that permits us to develop sound logic models and results chains. This is thanks primarily to the logframes that are found in GEF-3+ projects. This example is representative of other GEF-3+ projects that we reviewed in its level of detail, articulation of assumptions, and identification of intermediate results that are required to get to desired impacts.

Logic Model

Inputs	Outputs	Results		
		Short-term Outcomes	Intermediate Outcomes	Impacts
<ul style="list-style-type: none"> Financial resources Computer equipment and software Human resources 	Modernization of property registration system	<ul style="list-style-type: none"> Greater efficiency in processing land registrations More registered land Increased Municipal revenues from property taxes Increase staff of Municipal Environmental Units (MEUs) MEUs seek technical assistance from Sustainable Tourism Commission Municipal environmental bylaws implemented 	Reduction of: <ul style="list-style-type: none"> Habitat conversion and degradation due to residential development Water pollution from sewage Sedimentation 	Greater conservation of: <ul style="list-style-type: none"> coral reef systems beaches coastal rivers lagoons mangroves sea grass beds

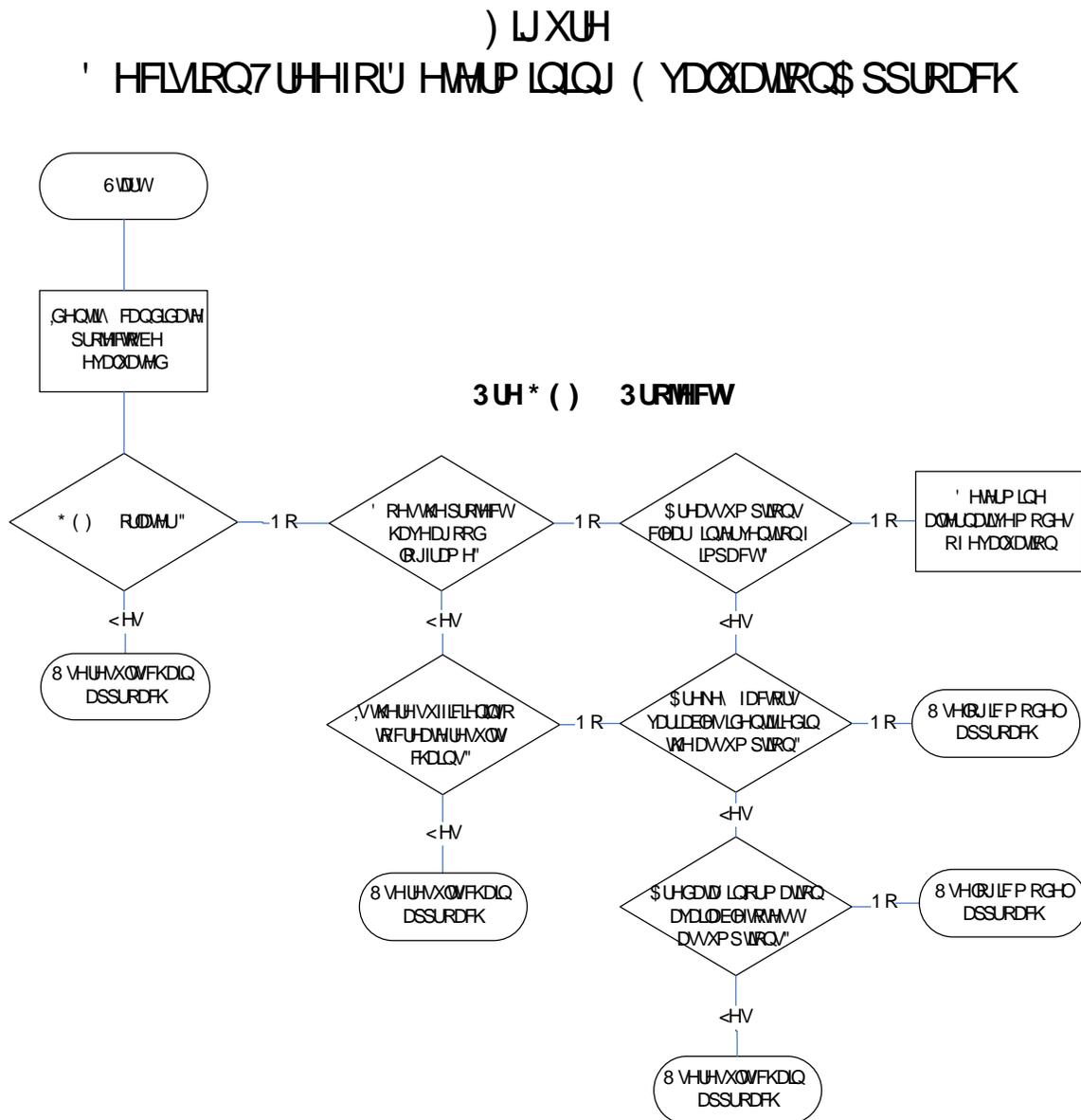
Results Chain



Summary of How to Apply a Theory of Change Approach

It is clear from the above findings that GEF-3+ projects lend themselves easily to being represented by both logic models and results chains. For many pre-GEF-3 projects as well, there are sufficient information and material to create results chains that can be used as the foundation of evaluation. And finally, for those pre-GEF-3 projects that do not lend themselves to a results chain approach, use of logic models to do a more general impact evaluation is possible.

Figure 6 is designed to assist the GEF EO team decide when it is appropriate to use the logic model approach and when it is appropriate to use the results chain approach.



Using a Theory of Change Approach to Aggregate Project Results

In addition to testing the assumption that a Theory of Change model would serve as an effective approach to evaluating GEF projects, we also wanted to see if it could be used to help promote cross-project and cross-program aggregation of results and to enhance learning within GEF. To carry out this aspect of our work, we analyzed the extent to which our sample of projects could be combined in various functional ways to maximize learning on particular themes. As we describe in the Methods section above, we used the model of a conservation project composed of actors, actions, threats/opportunities, and targets as the organizing structure for our analysis. And we used the IUCN/CMP classification of actions and threats to categorize and aggregate all actions and threats found in our sample of projects. By gauging the amount of overlap of key themes across projects, we hoped to determine if it was feasible to combine these projects for analytical purposes.

Looking at frequencies of threats, we found that, indeed there was substantial overlap. In the Table 1, for example, eighteen of the twenty projects we reviewed listed *Habitat conversion and degradation* as a major threat. Similar results were found for *Biological resource harvesting*. Simply organizing cross-project learning or aggregation of results on threats appears to be one viable option. Note that all projects address multiple threats so the denominator for each row is our sample size of twenty.

Table 1: Aggregation across Projects
Frequency of Common Threats (n=20 projects)

Habitat conversion & degradation	18
Biological resource harvesting	14
Pollution	4
Energy & mining	3
Transportation infrastructure	2
Invasive and problematic species and genes	2

When we look at aggregation across projects based on the strategies they employ, we find similar results. In Table 2, we include an analysis of strategy overlap for the highest order of strategies in the IUCN-CMP taxonomy. It is readily apparent that, at this macro level, there is much overlap in our sample of twenty projects. (As for threats above, the denominator for each row is our sample of twenty projects.)

Table 2: Aggregation across Projects
Frequency of Common Strategies (n=20 projects)

1. Land/water protection	7
2. Land/water/species management	20
3. Law & policy	12
4. Research & awareness	13
5. Economic & other incentives	3
6. Capacity building	13

Taking just one of our macro categories, Category 2 – *Land/water/species management* – for which all twenty projects in our sample had at least one strategy, we can further break it down into successively more refined strategies listed in the IUCN-CMP taxonomies. As demonstrated in Table 3, even at these more specific levels, we see there is ample opportunity for cross-project comparison, aggregation of results, and learning centered on project strategies.

**Table 3: Aggregation across Projects
Frequency of Common Strategies (n=20 projects)**

2. Land/Water/Species Management	20
2.1 Protected Area Management	15
2.2 Compatible Resource Use	11
- Forest management	5
- Sustainable resource use projects (general)	3
- Ecotourism	3
- Sustainable agriculture	3
- Water/watershed management	2
- Forest fire prevention and control	1
- Livestock management	1
- Medicinal and herbal plant conservation	1

Finally, we wanted to determine the extent to which – in our sample of twenty projects – we could functionally combine projects that mention the same combination of actions, threats, and targets. The probability of finding projects that overlap simultaneously on all three themes would seem to be low, yet in fact, Figure 7 illustrates that there is a significant amount of multi-theme overlap.

**Figure 7: Aggregation across Projects
Examples of Matching Combinations of Targets, Threats, Strategies (n=20)**

Conservation Trust Fund Strategy



Projects that have this identical combination of key themes include Integrated Biodiversity Protection in the Sarstun-Motagua Region (Guatemala), Conservation of Biodiversity and Sustainable Development in the Federal Islamic Republic of Comoros, Wildlife and Protected Areas Conservation (Lao), Protected Areas Program (Mexico), and National Trust Fund for Protected Areas (Peru).

Ecotourism Strategy



Projects that have this identical combination of key themes include Integrated Biodiversity Protection in the Sarstun-Motagua Region (Guatemala), Catalyzing Sustainability of the Wetland Protected Areas System in Belarusian Polesie through Increased Management Efficiency and Realigned Land use Practices (Belarus), and China.

Based on this finding, even from within a small sample, it is clear that a Theory of Change approach lends itself well to providing a structure to aggregate across GEF projects to document results at a greater scale and promote learning. However, there are a number of issues that arose during our analysis that will affect the Evaluation Office's ability to conduct effective Theory of Change evaluations across projects. These include:

- Inconsistent language, concepts, and definitions
- Inconsistent use and definitions of logframes
- Variability in the quality and availability of data
- Emphasis on measuring impacts (far right side of a results chain) and/or activities (far left side of a results chain) with little or no attention to the intermediate results – the boxes in the middle that would connect the logic between the activities and the anticipated impact.

Recommendations

The primary objective of this work is to determine the best way to evaluate the long-term results of GEF interventions. In particular, the EO wishes to assess projects a few years after GEF support has been concluded.

Based on the above analysis, it is clear that a Theory of Change approach offers great potential for satisfying the intentions of the EO. This confirms the significant amount of analysis and planning conducted by the EO in anticipation of this work.

Our recommendations in this section are organized into three categories:

- Vision for an ongoing evaluation program based on a Theory of Change approach;
- Necessary conditions to undertake a Theory of Change evaluation approach; and
- Piloting the Theory of Change evaluation approach.

Vision for an Ongoing Evaluation Program Based on a Theory of Change Approach

This section describes our suggestions for how a long-term impact evaluation program could be formalized in the EO. We have purposely called it a “vision” as we present it as a model for what we think the EO should pursue over the long-term – and we have done it without consideration of budget or staffing. If indeed, this is consistent with the vision that the EO wishes to pursue, then financial and staffing adjustments will need to be considered and brought before Council. This section is meant to serve as the broad context within which the piloting activity – described below in more detail (including a budget and timeline) – is placed.

We divide the recommendations related to this broad vision into three primary categories: A) Overarching recommendations related to the vision; B) Evaluation of individual projects; and C) Evaluation of themes across projects. We also provide recommendations for the conditions necessary to undertake a Theory of Change evaluation approach. For reference purposes, we have numbered the recommendations. They are not, however, numbered according to any order of priority.

A) Overarching Recommendations Related to the Vision

While these may seem obvious, we wish to state them here for the sake of clarity and completeness. In subsequent sections of this report, we include details on how to implement some of the general recommendations listed here.

Rec. 1. Adopt a Theory of Change approach for all future evaluations conducted by the EO

While restating what we have said before, the considerable work done by the EO prior to this work and our own analysis confirm that a Theory of Change approach has great potential as an evaluation process for GEF. While the EO has committed to conducting a pilot phase of whatever approach is selected for future evaluations, we believe that the

Theory of Change approach is the one that should be adopted for all evaluation work in the future for GEF.

Rec. 2. Standardize methods and indicators in the context of a Theory of Change approach to the extent possible

For a Theory of Change approach to be effective for the GEF portfolio, it must be standardized in a number of areas. First of all, the way the evaluations are conducted must be standardized. This means that, regardless of who conducts evaluations, they are done using the same principles, concepts, and tools. We recommend using the two primary methods described in the Background section of this report – namely logic models and results chains. All evaluators should apply these methods consistently across all evaluated GEF projects. Furthermore, evaluators must rely on secondary data and project reports, interviews with key personnel, and the collection of primary data in a consistent format across projects.

Second, the variables that are used to conduct the evaluations should be as consistent as possible. While in a subsequent section we go into more detail about the standardization of the information that is collected and used to evaluate GEF projects, in general terms, the more consistent variables can be in and across individual evaluations, the higher the likelihood that results can be aggregated from the project level to the program or thematic levels. By allowing for this level of aggregation, the EO will be better positioned to measure success across the GEF portfolio and generate more generalizable, yet non-trivial lessons that can be incorporated back into individual ongoing projects and future GEF-funded actions.

Rec. 3. Consider integrating a Theory of Change approach into the design phase of GEF-funded projects

In recent years, GEF has taken great strides to more explicitly identify assumptions – and ways to measure those – at the beginning of the design phase. This was immediately apparent during our analysis of our sample of projects, with the GEF-3 projects having the benefit of logframes. This work should continue to mature and be bolstered by the EO's adoption of Theory of Change evaluations. In particular, by working with program managers who are responsible for overseeing the design of new projects, the EO can help planners explicitly link interventions to expected outcomes and impacts and determine the variables most likely to be used to test underlying assumptions. In this way, the EO will be helping project managers to practice adaptive management by monitoring these key variables during the life of the project, but also, it will produce concrete data that can be used in any post-project terminal evaluation. As such, the EO will be facilitating its own implementation of future evaluations by having input into the variables it wishes to see collected in order to evaluate impact. In addition, the generation of data related to key variables during routine project monitoring can be used by evaluators during and after project implementation.

We recommend that the EO either adopt or endorse the emerging consensus in the conservation community on project design, implementation, and monitoring/evaluation as reflected by CMP's *Open Standards for the Practice of Conservation*. This framework is based on a Theory of Change approach and is thus completely compatible with the

findings and recommendations of this report. In fact, the *Open Standards* provide the context within which a Theory of Change approach to evaluation is most appropriate. By using the *Open Standards*, therefore, the EO can more simply and directly demonstrate to GEF and IA staff how the evaluation fits with other components of project execution.

Rec. 4. Link the evolution of the GEF EO Theory of Change evaluation approach to ongoing work in the conservation community

One problem that seems to be present in conservation work around the world is that the various institutions that implement or otherwise support field efforts – including NGOs, bilateral and multilateral agencies, and donors – spend relatively little effort to ensure a flow of learning across institutional boundaries. With respect to the work being conducted by the EO, it has produced much analysis and thinking that would be of great service to the rest of the conservation community. Likewise, in recent years, there have been significant innovation and advances made by other conservation organizations that have direct bearing on the work the GEF EO is planning to undertake. For example, in addition to the *Open Standards* mentioned above, CMP has produced or is the process of producing an approach to conservation audits, software for indicator selection and project management, and a Rosetta stone, among other relevant products, for the conservation community. Below, we describe some of the CMP products in more detail.

Rec. 5. Use the Theory of Change evaluation approach as an opportunity to build the capacity of GEF and Implementing Agency staff

While the EO could fulfill its mission simply by conducting relatively independent evaluations of specific projects, it would be missing out on an important opportunity for the GEF family. In particular, the EO could involve GEF and IA staff in the design and implementation of evaluations, using the Theory of Change approach. By doing this, the EO would be contributing to staff development and the capacity of the GEF network to better conceptualize, design, implement, monitor, and evaluate its conservation actions. At present, except in isolated cases including, in particular, the biodiversity focal area, there is little evidence that GEF projects benefit from the clear and analytical thinking at the conceptualization phase that is required to fully embrace a Theory of Change approach. Often, at best, assumptions are implicit rather than explicit, and the linkages between actors, action, outputs, outcomes, and impacts are not clear. By more closely involving GEF and IA staff in future evaluations, the EO will be contributing to the capacity of GEF staff to better position their projects for improved execution and evaluation.

B) Evaluation of Individual Projects (For IAs)

Although the evaluation of specific projects is the purview of individual IAs, the EO can design its evaluation approach so as to help IAs assess project impact in the six categories listed above in the Background section. These include direct impact, sociopolitical sustainability, programmatic sustainability, institutional sustainability, financial sustainability, and replication. In the following section on piloting the approach, we provide more detail on how to go about assessing these categories.

The focus for IA-led project-specific evaluation should be to determine what worked, what seemed not to work, and why with emphasis on developing lessons-learned. As such, the results of individual project evaluation should include specific findings directed in very practical terms to project managers and country representatives responsible for post-project activities. In addition, specific, generalizable lessons from individual project evaluations can be used to report to and inform other entities, (such as GEF managers and the Council) in order to facilitate repeating successes and avoiding failures in the future.

While the main purpose of this work was to develop an *impact* evaluation approach for the EO, we believe that it is almost just as important for the EO to promote a framework for the evaluation of the *process* that guides project implementation, including the underlying logic, structure, and function of GEF-funded projects, in order to fully understand why they succeed or fail. In order to do this, we recommend developing or adopting an audit approach to complement the impact evaluation protocol. For this reason, we include a specific recommendation related to project auditing.

Key recommendations for the evaluation of projects included in GEF EO studies:

Rec. 6. Base evaluation on theories of change used or implied by the specific project

While this may seem obvious, sometimes evaluators evaluate projects based on what they believe the project *should have* been trying to achieve instead of what they were designed to achieve. This sometimes happens in projects for which expected results are not clear from early project documentation.

To avoid this situation, GEF evaluators should invest a considerable amount of time reviewing relevant documents and interviewing informed staff. In some cases, evaluators will be able to quickly reconstruct theories of change from existing project documentation. In other cases, evaluators will have to sit down with individuals who were a part of the design or implementation of the project in order to reconstruct the implicit theories of change that linked project activities to expected impacts.

Rec. 7. Use best available standardized chains and indicators

Instead of reinventing chains and indicator sets for every evaluation done by the EO, much efficiency can be gained by developing a library of theories of change chains and the indicators that can be used to test and evaluate them. While some people may argue that all projects are unique and therefore coming up with common chains and indicators is not possible, based on our experience, we would not agree with this conclusion. In fact, it is surprising how similar projects may be, and the underlying assumptions – represented by theories of change – may be remarkably similar in projects that take place in seemingly different settings. What fundamentally vary from site to site are the context variables where projects take place.

At the same time, if similar chains can be used across multiple projects and thus similar results are expected, then it is highly likely that similar indicators – or families of indicators – can be used to evaluate similar theories of change across these sites. Again, this would represent increased efficiency as evaluators would not have to spend time

searching for relevant indicators or otherwise developing new indicators for their respective evaluations.

Rec. 8. Push evaluation to the impact side of the theories of change to the extent possible

While attention should be paid to the activities and inputs side of theories of change in the course of conducting evaluations, to the extent possible, evaluators should always strive to gauge impacts (found on the far right of Figure 2). If it proves impossible to assess impact – either through secondary or primary data – then evaluators should rely on outcomes data to determine the extent to which a project succeeded. This should be relatively easy to accomplish taking a Theory of Change approach as outcomes and impacts should be clearly represented and thus evaluators should know what to analyze before they initiate evaluation activities. Thus, if they encounter no opportunity to collect impact-related data, then they will, at least, be prepared to examine the results of the various outcomes associated with the project.

Rec. 9. Standardize approach procedures across evaluators

The EO should not assume that all evaluators will conduct Theory of Change evaluations in the same way. It is therefore essential that the EO develop and institute some form of training or briefing that shares with potential evaluators the approach it wishes to use. In particular, evaluators need to know how to create logic models and results chains that fulfill the requirements of the EO. And they need to know about criteria for indicator development adhered to by the Office. To the extent possible, all evaluators should be conducting evaluations in the same manner, using the same tools, and identifying indicators in a similar fashion across all evaluations.

Rec. 10. Adopt or develop an audit approach that analyzes process

Often, project success is determined not by the extent to which an intervention or strategy is based on sound logic linking the intervention to a desired impact. Rather, it might be a function of more mundane factors related to project administration, management, or implementation. To be able to test whether such factors are influencing the success of GEF-funded projects, we recommend that the EO develop or adopt a project auditing tool that specifically addresses process factors (i.e., how a project was conceived and carried out), not impact.

C) Evaluation of Themes across Multiple Projects

The greatest potential for the EO lies in its position within GEF to look at higher-order institutional results and to facilitate accelerated learning across projects. The universe of GEF-funded projects is quite large and the potential for strategically dividing it up into a series of learning themes is equally extensive. Because of the many projects that are supported by GEF and the range of conditions in which they are implemented, the power of these thematic-focused samples can be considerable.

Using the diagram of a generic conservation project that we included in Figure 1, it is clear that the EO could organize systematic learning on a number of categories of themes. The most obvious way of organizing learning could be around the four main columns:

targets, threats (or opportunities), actions, or actors. Choosing which category to focus on will depend on the question that the EO wishes to address.

For example, if the EO wanted to gather lessons-learned regarding GEF's experience funding projects in a marine setting, it might select all projects that are designed to conserve coral reefs (a target) around the world and determine what has worked and what has not worked in this particular sample. Of course, in this case, threats, opportunities, and strategies will undoubtedly vary greatly across sites. For this reason, it may be advantageous to further refine the sample based on these factors.

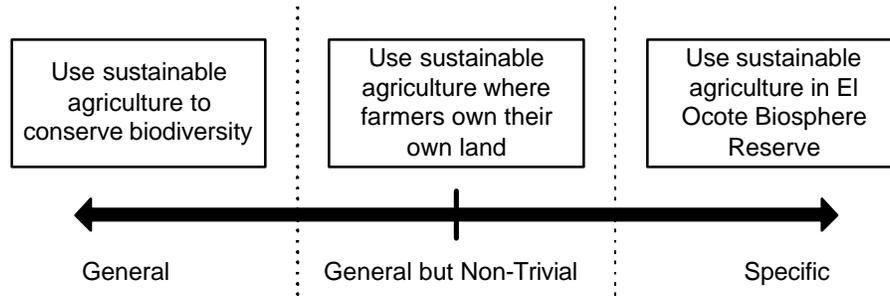
Similarly, the EO might wish to look at GEF's experience dealing with invasive species and thus organize an evaluation sample around this particular threat. In this case, targets and actions will vary, so again, further refinement of the sample may be necessary. But what would come out of this evaluation are lessons-learned related to the extent to which GEF-funded projects led to the abatement of a particular threat.

Finally, and perhaps more functionally, the EO might wish to focus on analyzing the efficacy of a particular strategy that GEF funded in the past or is currently funding. For example, the EO might wish to assess GEF experience implementing policy reform in a suite of projects across the portfolio. Here too, it would be wise to limit analysis to specific threats and targets as there will be wide variance in these across the entire GEF portfolio.

By planning, conducting, and pooling evaluations in a way that the EO can look across similar projects working in similar enough conditions, general and yet non-trivial principles can be derived. This approach involves controlling for key factors such as targets, threats, and actions while looking for and embracing the variance in other factors and conditions that we would expect to see from site to site.

By taking this approach, the EO would not be attempting to conclude that a given conservation action works or does not work under all conditions. Instead it would be attempting to arrive at much more subtle, yet significant conclusions. Ideally, at the end of a cross-project evaluation that is organized around a specific theme, the EO would be in a position to say: "Strategy X achieves intended conservation outcomes and/or impacts under the following conditions:..." (and, conversely, "It will not work under the following conditions:..."). This kind of guidance is of most use to project managers – at all levels – who are trying to optimize their work in the field. It would also serve GEF and the IAs to help them select the right strategies for future projects and avoid those that are not appropriate. In this way, the EO would be in a position to share *general and yet non-trivial principles* both internally and externally.

The best principles are those that apply at a wide range of sites, but are not so trivial that practitioners will disregard them. As shown in the right side of the following diagram, at any given site there are *specific* principles that are of great use to people working at that site.

Figure 8: General and Yet Non-Trivial Principles

For example, project team members working at a site in Mexico might develop a principle such as:

P – Use sustainable agriculture in El Ocote Biosphere Reserve

Unfortunately, these site-specific principles do not really help a person working at the next site over, let alone at a site halfway around the world. On the far-left side of the diagram are *general* principles that apply to most or all sites as illustrated by the example:

P – Use sustainable agriculture to conserve biodiversity.

Unfortunately, these principles tend to be trivial – they may be true but are not very helpful to practitioners. Are there *general and yet non-trivial* guiding principles as shown in the center of the diagram? It is most likely that, if these guiding principles exist, they will take the form of conditional probability statements. For example, we might develop the principle:

P – Use sustainable agriculture where farmers own their own land.

This principle applies to more than one place but not everywhere. Furthermore, it refers to a specific strategy to undertake (use sustainable agriculture), outlines specific conditions under which the principle will hold (when farmers own their own land), and leads to specific and measurable impacts (improved conservation). This principle is not guaranteed to work in all instances, so the user must be smart enough to apply it to his or her own situation. The task thus becomes determining not just what the principles are, but also under what conditions and with what probability of success is each principle likely to work.

Unfortunately, far too many of the “lessons learned” in conservation are either too general (e.g., “involve local people in projects”) or too specific to be of much use. As a general rule, topics that practitioners are mostly likely to be interested in are solutions for dealing with specific problems they are facing. In conservation parlance, this involves strategies for dealing with threats to biodiversity or taking advantages of opportunities that will make their work more effective.

Specific recommendations for evaluating themes across multiple projects include the following:

Rec. 11. Strategically select a sample of projects that allow the EO to effectively compare across sites

In order to aggregate results and create general and yet non-trivial principles as we described above, it is essential that the EO select a sample – or samples – of projects that are organized in a strategic fashion. As described above, this means selecting a sample based on biodiversity targets (intended impacts), threats or opportunities, or actions – or some combination of these. We recommend selecting projects to be evaluated based on these themes rather than by region or IA, although we realize that the EO will want to balance regional and IA representation to the extent possible.

Rec. 12. To the extent possible, limit the number of theories of change to be used in thematic evaluations

Clearly, there are many, many theories of change that could be created for a set of conservation projects. Given the limited resources that the EO will undoubtedly have to conduct evaluations, we recommend that it work with relevant GEF staff to determine the areas or themes of greatest interest. While the primary focus should be on a limited number of chains if possible, this does not preclude the EO from doing any other evaluations – still based on a Theory of Change approach – not directly related to focal chains.

Rec. 13. Rely on input from thematic experts to help create theories of change and identify appropriate indicators

If theories of change do not already exist in the files of the EO from previous evaluations, then the Office should consult with thematic experts to help develop appropriate logic models or results chains. Many of these experts will no doubt be available from the ranks of GEF and IA staff. At times, however, the EO may wish to consult with outside experts. Starting with expert-based theories of change will accomplish three things: It will help the EO base its work on the conventional wisdom in the field – starting with what is already known; it will make the process of determining which variables to use in the evaluation a more efficient process; and it will deflect any potential criticism of conducting evaluations not solidly founded in theory and knowledge.

Necessary Conditions to Undertake a Theory of Change Evaluation Approach

A major benefit of a Theory of Change approach to evaluation is that the EO will be able to clearly trace assumed linkages from interventions to outcomes and impacts and determine the extent to which projects achieved what they intended to achieve in the manner they intended to do it. The true power of this approach, however, lies in its ability to look at results across projects, to figure out what works and what doesn't work under different conditions, and to aggregate results across portfolios of projects.

In order to fulfill the potential of this approach, however, the EO must meet some fundamental conditions. These relate primarily to the standardization of process and components of a Theory of Change approach. The following recommendations rely

heavily on products developed by CMP as this consortium of conservation organizations came about as a direct result of perceived need to communicate better across institutions, and thus, the products that are generated by CMP are designed to accomplish this.

Rec. 14. To the extent possible, use a standard taxonomy of components of a Theory of Change approach

For a Theory of Change approach to work, evaluators must be able to quickly illustrate linkages between actions and outcomes/impacts in a way that is generally understandable to the conservation community and translatable across projects. To facilitate communication, it is therefore advantageous to have evaluators use a common nomenclature. This is true, in particular, for the main components of a Theory of Change approach as described above: actions, threats/opportunities, and targets. The more unified and consistent the definitions of these components are across evaluations, the more comparable they will be. Rather than develop its own taxonomy, the GEF EO should consider using the IUCN-CMP classifications found at the CMP website (www.conservationmeasures.org).

In addition to the classifications of the components of theories of change, cross project comparison and aggregation of results would be enhanced by consistently using terms and concepts in project design, implementation, monitoring, and evaluation. As noted in the Findings section above, GEF and the three IAs use terms and concepts in inconsistent ways across institutional boundaries. In order to overcome this – or at least have a common language that can be translated across institutions – the EO should adopt existing “industry standards” for concepts and definitions.

At present, there appears to be two main sources for the harmonization of terms and concepts. The Organization for Economic Co-operation and Development has created the DAC Glossary, which can be found through the OECD website (www.oecd.org). This extensive publication lists a wide variety of terms and concepts and is based on prevailing evaluation field standards. The publication, however, does not include many of the terms in common use in the field of environmental conservation.

An addition to the DAC glossary is recent work conducted by CMP, including the CMP Rosetta stone that is based on analysis of concepts and terms from all member organizations. The Rosetta stone not only shows where different terms mean the same thing across organizations, but it also provides organization specific definitions and a suggested global definition for these key terms. The Rosetta stone can be found at CMP’s website (www.conservationmeasures.org).

Rec. 15. Use similar variables and indicators across evaluations

In addition to using the same terms and concepts, it is vital that evaluators use the same metrics to measure impact and changes in key factors along theories of change chains or logic models – especially if the EO wishes to compare results across projects or roll results up to gauge higher levels of impact. While there are endless lists of indicators available from multiple sources, these lists tend to focus primarily on biodiversity targets and, to a lesser extent, on threats to biodiversity. In order to evaluate conservation

interventions using a Theory of Change approach, it is important to analyze the path of association from intervention to impact. To do this, evaluators must have access to indicators related to variables that are found in the inputs, outputs, and outcomes portions of the chains as well. CMP is currently supporting work to develop the Strategic Indicator Selection System (StratISS) and the EO may wish to tap into this resource as well.

Rec. 16. Adopt a common framework within which GEF evaluations will take place

As mentioned in the Background section, the Theory of Change approach that seems most appropriate to the GEF context is a combination of logic models and results chains. Logic models should be used only when there is insufficient documentation to create results chains.

While GEF EO's mandate does not include performance at the project level, it is worth mentioning that a additional benefit to a theory of change approach is that it provides project managers with a way of incorporating results into decision-making. In order to do this, evaluations should be carried out in the context of a framework of project design, implementation and evaluation. While in recent years, most conservation organizations have gone towards project cycle management, it is only recently that this concept was more tailor-made to the conservation community. CMP has adapted traditional project cycle concepts to conservation and produced a document that provides much more detail than was previously available to help operationalize the standards. The product is *The Open Standards for the Practice of Conservation*, which relies on a Theory of Change approach as the conceptual framework. The most recent version of the Standards is available at the Conservation Measures Partnership's website:

<http://conservationmeasures.org>

Conclusions

Based on the analysis of the ten Pre-GEF-3 and ten GEF-3 projects, a Theory of Change approach to impact evaluation seems like the best way forward in the GEF portfolio. This approach provides a roadmap for analyzing how and why an initiative works. It delineates the pathway of an initiative by making explicit both the expected results of the initiative (early, intermediate, and long term) and the actions or strategies that will lead to the achievement of results.

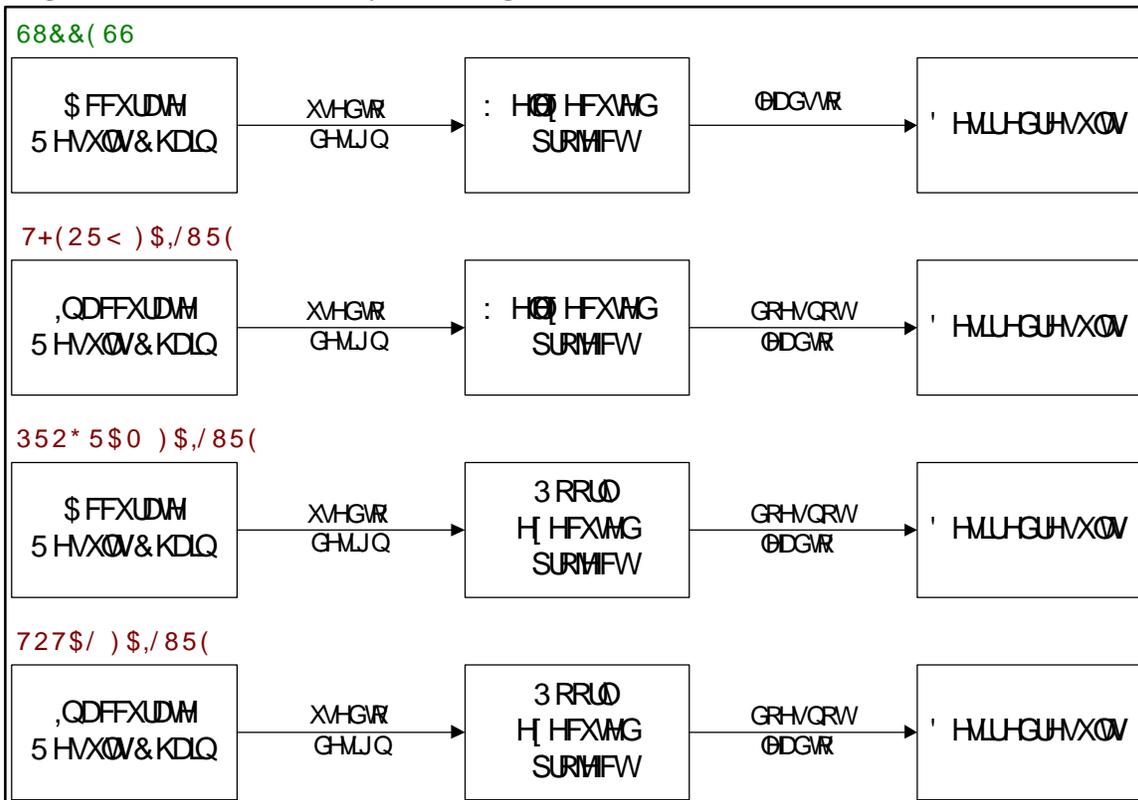
In the purposeful sample we reviewed for this work, we found that there is fairly stark difference in terms of logic and clarity between pre-GEF-3 projects and GEF-3 projects with the newer projects significantly better articulated than the old ones. Of the older GEF projects, however, about half include enough desired project results/outcomes to be able to construct at least logic models and simple results chains.

In order to accommodate the diversity of pre-GEF-3 and GEF-3+ projects, we have found that there must be at least two levels of the Theory of Change approach. The first approach must be able to accommodate those projects for which the theory of change is

not very explicit, assumptions are not articulated, key variables for impact evaluation are not identified, and data on key indicators are not collected. In order to evaluate projects under these conditions, we recommend taking a Theory of Change approach that is represented by *logic models*.

For those GEF projects that do have well articulated interventions and expected results – and the assumptions that link the two – we advocate a more rigorous form of the Theory of Change approach. This approach is best represented by *results chains*. Applying the more rigorous results chain approach allows evaluators to determine the extent to which projects have been successful. In particular, as illustrated in Figure 9, it facilitates analysis of not only the extent to which a given intervention worked, but also why it worked or did not work.

Figure 9: Causes of Theory and Program Success and Failure



There are some challenges that will have to be overcome if a Theory of Change approach is to be successful. While these were readily apparent from our analysis, they have also been mentioned – often multiple times – in previous evaluations of GEF work. In very brief terms, these include the following:

Absence of agreed-upon terms and concepts – In our review, we found a broad range of terms and concepts being used by GEF and the IAs. While on the surface this may seem like a trivial point, it can, indeed, have serious consequences. This is true especially when organizing logic models and results chains. In the projects we reviewed for this work, different managers called the same concept an indicator, objective, output, or outcome. In order to truly measure impact, and, most importantly, compare across projects, this issue must be resolved. GEF should look to other communities and fields to adapt and adopt common terminology as there are widely accepted terms already in play in evaluation circles.

No one unified taxonomy of actions, threats, and targets – While some previous evaluations of GEF projects and programs arrive at the conclusion that all projects are different and therefore, cannot be compared, we do not share this view. In our experience, there is a finite set of conservation targets, threats, and actions that practitioners use in order to achieve conservation results. Furthermore, there are finite ways in which these elements combine into logic models and results chains. In order to harmonize

evaluations, capture lessons, and systematically build a knowledge based about what works, what does not and why, GEF must employ an evolving and expanding taxonomy.

Lack of common logic to create common chains – In order to build a knowledge base in GEF, evaluators should use try to harmonize emerging logic models and results chains to the extent possible. By replicating these across time and space, GEF will learn more quickly and increase effectiveness and efficiency more rapidly.

Limited systematic structure and process to perform project evaluations – Another situation that struck us when we reviewed pre-GEF-3 Terminal Evaluations was the extent to which evaluation methods, approaches, and styles varied across these evaluations. Again, in order to be comparable and aggregate results across project to roll up to GEF-wide impacts, the evaluations themselves must be done in a consistent and standardized manner. In addition, data and inputs should be recorded in a systematic and consistent fashion.

Despite these challenges, we are convinced that GEF is well-positioned to employ a Theory of Change based approach to evaluation. GEF is at a critical point in its evolution and we believe this approach will help it better understand what has worked and what has not worked in the past, and what it can do to improve the way that it achieves and documents results, learns, and adapts.

GEF IMPACT EVALUATIONS

Piloting the Theory of Change Evaluation Approach

Draft Terms of Reference

1. Background and Rationale

The GEF portfolio has now matured to a stage at which consideration can be given to the long-term impacts of its projects. With this in mind, in 2006, the GEF Evaluation Office (EO) funded a study to identify an effective approach to measure the impacts of GEF-funded activities. The study recommended that the GEF adopt a Theory of Change approach and provided guidance for designing an evaluation approach based on the Theory of Change model. This approach provides a roadmap for analyzing how and why an initiative works. It delineates the pathway of an initiative by making explicit both the expected results of the initiative (early, intermediate, and long term) and the actions or strategies that will lead to the achievement of results.

Theory based evaluation adopts a logical model of cause-effect linkages, through which an intervention intends to achieve its objectives and explores empirically the extent to which events followed the anticipated sequence. The evaluation assesses the validity of the theory of change adopted (implicitly or explicitly) by the intervention, as well as the interaction between the specific local circumstances and the general principles expected to generate the desired change. The approach enables detailed examination of the nature of linkages between a complex set of causes and a set of effects. The theoretical model can clarify the assumptions upon which the intervention was based and also incorporate opposing theories, which may be applied to the same situation.

Adopting a Theory of Change approach will provide GEF with a mechanism to look at its results across its portfolio of projects instead of looking at results exclusively in terms of individual projects. It will help GEF better understand what has worked and what has not worked in the past. Finally, a Theory of Change approach to evaluation will facilitate GEF's efforts to enhance its impacts by documenting results, accelerating learning, and providing a mechanism for systematic adaptation.

2. Purpose

The main purpose of the pilot implementation of the Theory of Change approach is to test this approach on GEF's biodiversity focal area, using both logic models and results chains to lay out and test assumptions linking project interventions to desired biodiversity impacts. In particular, the pilot will develop the necessary instruments and protocols to perform evaluations in the future. As such, evaluators working on the pilot should develop draft theories of change for each project included in the study based on existing literature, expert advice, and project documentation. Evaluators will need to identify the right people to be involved, including key GEF and IA staff that will help set up field visits and other interviews. And they will have to talk to technical staff GEF and the IAs

to help refine and complete the draft theories of change upon which the evaluation will be based.

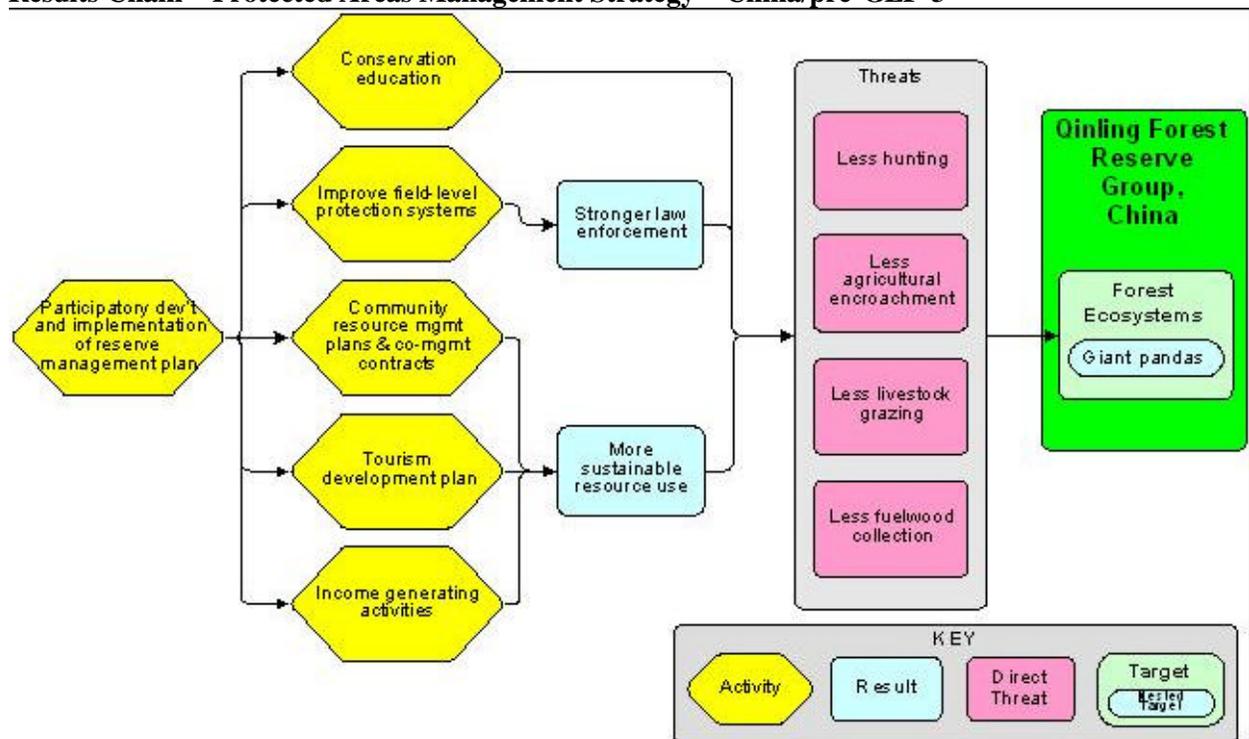
Illustrative Case

For illustrative purposes, the Terms of Reference draws on information and data from a pre-GEF-3 protected areas management strategy in China (*Nature Reserves Management*). Presented below are a logic model and results chain, drawn from the results of the 2006 GEF Impact Evaluation Study. The logic model does an adequate job of describing – in fairly general terms – the relationship between inputs, outputs, outcomes and impacts. The results chain, however, should ideally provide more detail about necessary intermediate results that lead to a desired outcome. In this case, most of the factors in the chain are either actions or threats, with little understanding of the conceptual or programmatic linkages that connect them. Nevertheless, this protected areas management strategy serves as a useful example for framing some of the issues the pilot study should explore.

Logic Model – Protected Areas Management Strategy – China/pre-GEF-3

Inputs	Outputs	Results		
		Short-term Outcomes	Intermediate Outcomes	Impacts
<ul style="list-style-type: none"> • Financial resources • Human resources • Equipment 	<ul style="list-style-type: none"> • Participatory preparation and implementation of reserve and land use management plans • Improving field level protection systems • Development of community resource management plans and co-management contracts between communities and reserves • Development of income generating activities for communities • Preparation of tourism development plans • Conservation education 	<ul style="list-style-type: none"> • stronger law enforcement • more sustainable resource use 	<ul style="list-style-type: none"> • Less agricultural encroachment • Less livestock grazing • Less hunting • Less fuelwood collection 	<ul style="list-style-type: none"> • Greater conservation of forest ecosystems in reserves and in biological corridor linking reserves • Greater conservation of giant panda habitat and giant panda population

Results Chain – Protected Areas Management Strategy – China/pre-GEF-3



3. Scope

The pilot phase of the Theory of Change approach will focus on the biodiversity portfolio’s protected areas strategy. The evaluators, however, are expected to work with the EO to clarify the specific actions should be included within the broad category of protected areas (for example, enforcement, park guard training, environmental education, and community management).

The EO has chosen the protected areas (PA) strategy as the focal theme for the pilot phase for a number of reasons, including the following:

- PAs are one of the primary approaches supported by GEF biodiversity focal area and the IAs in the past;
- PAs are the first strategic priority of the biodiversity focal area, and as such, are expected to remain a pillar of GEF biodiversity funding in the future;
- The 2004 biodiversity program study specifically noted that an evaluation of the GEF support for protected areas has not been carried out, and specifically recommended that such a study be undertaken;
- Protected areas are an obvious change theory, not just for the GEF, but for conservation as a whole;
- In many areas, substantial field research has been undertaken, and of all the places where GEF interventions are occurring, protected areas are likely to have the most information in terms of baselines and biological data;

- A PA strategy could be addressed at both a thematic AND regional cluster level (e.g. protected areas in southern Africa, or Central America, SE Asia, etc.);
- The protected areas concept is more clear from a theoretical point of view than some of the other concepts in biodiversity conservation; and
- The biodiversity focal area team has made great strides in identifying appropriate indicators for protected areas through its “Managing for Results” system for Strategic Priority One – Catalyzing Sustainability of Protected Area Systems.

In general, as the purpose of the pilot is to test the feasibility of the approach, it should be conducted incorporating projects for which the EO perceives a high likelihood that theories of change can be constructed and tested. As such, the EO should not be too worried about the sample included in the pilot at this stage and should try to include these “low hanging fruit”. If it works in these, then it might work elsewhere in more complex and difficult situations. If it doesn’t work in the simplest of projects, it is surely not going to work in more difficult projects. But if the approach is tried on projects that are too complex before being fully tested, there is a higher probability that the test will fail, not because of some inherent flaw in the methodology, but instead because of other intrinsic factors or challenges in the project itself for which we cannot control.

Key Questions

In its *Final Draft Approach Paper*, the EO identified some key questions that should be addressed during the pilot phase. These questions should help drive the collection of relevant data to at once test the approach and refine it so as to prepare for more full scale implementation of the impact evaluation approach with other projects

In addition to gauging the extent to which project implementers were able to achieve results down the theories of change pathways (the impact aspect of the evaluation), the pilot team should develop the data collection instruments to capture the analysis of five additional categories – sociopolitical sustainability, programmatic sustainability, institutional sustainability, financial sustainability, and replication. Examples of possible questions that could be included for each category include:

Impact evaluation

- What evidence exists that documents a change in indicators related to the intended impact of this project?
- What evidence exists that can demonstrate an association between the intervention that was funded and the impact? For example, in the China Protected Areas Strategy project, how has tourism development affected giant panda populations? Are there intermediate results that show a clear path from action to impact? For example, has tourism development led to more sustainable resource use? Has this sustainable resource use, in turn, led to less agricultural encroachment? Has the decrease in agricultural encroachment maintained or increased panda habitat? Have there been any changes in panda populations?
- If impact cannot be measured, are there outcome measures that can serve as proxies? If so, what are they? For example, will hunting levels and the rate of agricultural

encroachment give the GEF EO a fairly good indication of the likely health of the giant pandas?

Sociopolitical sustainability

- Is there any evidence that the intervention supported by GEF funds directly benefits key stakeholders? If so, who benefits and how? For example, who participates in the tourism industry locally? How do local businesses benefit? How do local Chinese villages and individuals benefit?
- Have linkages been established with key government departments and individuals to ensure the project continues after GEF funding?
- Is there any evidence that sectors of society feel ownership over project activities and will want to perpetuate them?

Programmatic sustainability

- Are programmatic actions that were initiated during the project continuing after GEF funding ended? For example, has the Chinese government made commitments to maintain or increase field level protection systems?
- Have other donors, agencies, or organizations stepped in to provide resources once GEF funding ended?
- Have communities or other key individuals picked up any projects activities to perpetuate them after GEF funding ceased?

Institutional sustainability

- Has the project led to any improved capacity within partner organizations that participated in the project? If so, what evidence is there? Who are the other partner organizations, and what role have they assumed in protected areas management since the start of the project?
- Have any local or national project partners gained in prominence in the country as a result of the project? If so, what evidence is there?
- Have management or governance structures improved in any participating partner organization as a result of the project?

Financial sustainability

- Have any of the project activities resulted in generating sufficient financial resources so as to offset some of their costs? For example, has tourism development increased park fees or contributions to directly support the PA?
- Have any participating project partners been able to leverage GEF funds for longer-term institutional or program funding? For example, has a trust fund been established to help fund protected area management in the future?
- Are additional financial resources available from other sources for project-related funding activities that otherwise would not have been there without the initial GF grant.

Replication

- Is there any evidence that project partners or managers are modeling other programmatic activities on this project? For example, are other protected areas in China using educational materials developed through this project to encourage citizens to work to protect other threatened species?
- Is there evidence that lessons have been documented in a way that project activities can be conducted and improved in other project sites?

Projects to Be Included

The pilot of the Theory of Change approach will focus on three projects. The Evaluation Team will determine the criteria for final selection of the projects but they should focus primarily on protected area management, and where possible, should have followed a Theory of Change-related approach (such as using logframes) and have readily available data. In addition, the EO wishes to balance selection of the sample of three across regions – for example, Africa, Asia, and Latin America.

Although it would be ideal to have representation of each of the three IAs, for the pilot phase, the pilot should not be constrained by this. In particular, since the types of projects that are funded differ across IAs, they may not be similar enough to test the desire to aggregate results up to a higher level. As recommended in the 2006 GEF Impact Evaluation Study, the sample should be strategically selected so that the EO can effectively compare across sites (Recommendation 11). For the purposes of this pilot test with such a small sample, it is better to select projects that are as similar to each other as possible. If it is politically feasible, in fact, and the IAs are willing, then the Evaluation Team should select projects from one of the IAs to conduct the pilot – the pilot should decrease variability and unforeseen differences across projects to the extent possible.

4. Methodology

Approach

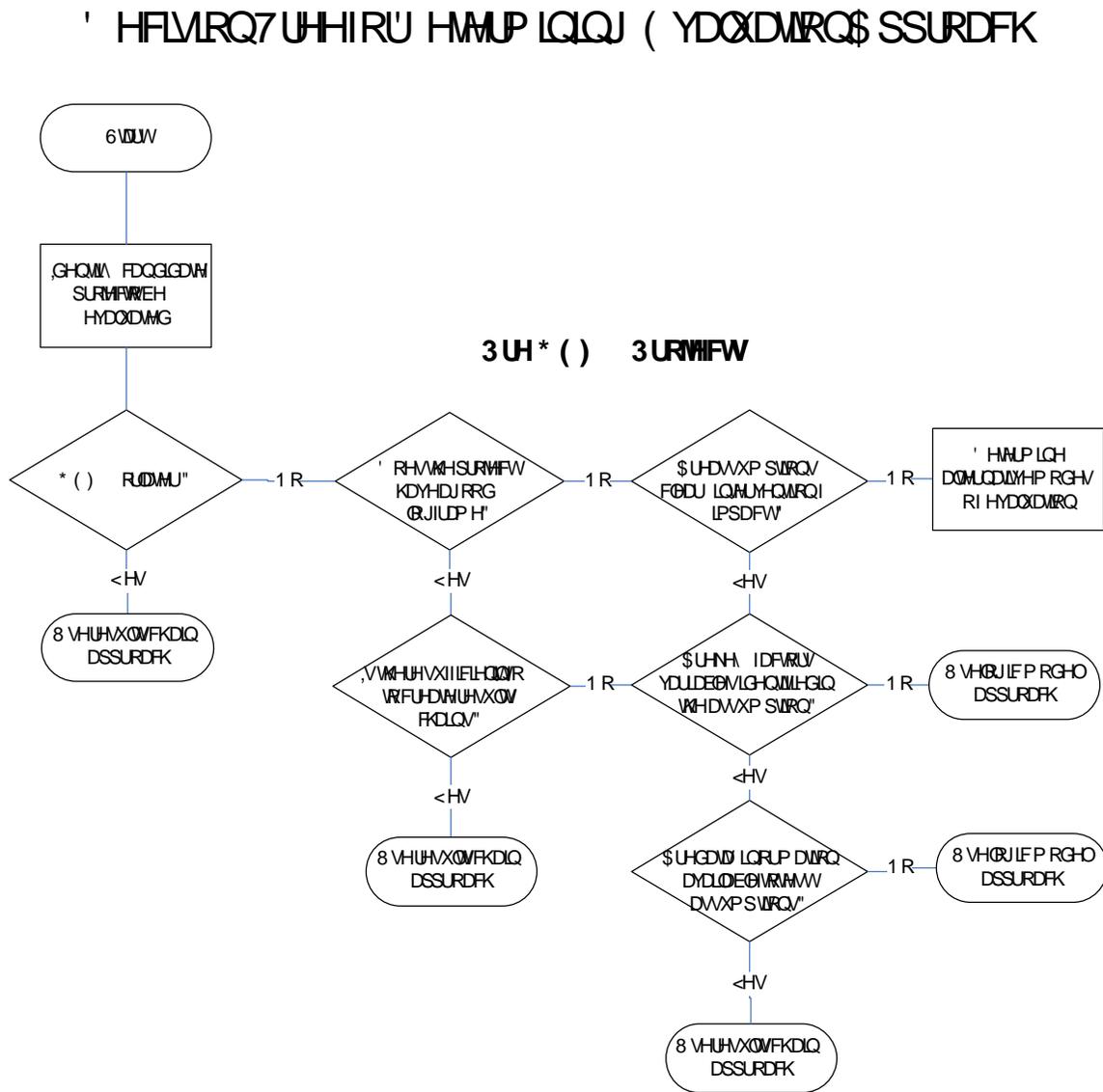
As recommended in the 2006 GEF Impact Evaluation Study, the pilot evaluation – and presumably, subsequent evaluations of GEF-funded projects – should be based on the theories of change used or implied by the specific project (Recommendation 6).

Evaluator should use the best available standardized chains and indicators (Recommendation 7) to avoid “reinventing the wheel,” but also to facilitate learning across projects, as well as rolling up data across projects.

In order to accommodate the diversity of pre-GEF-3 and GEF-3+ projects, there must be at least two levels of the Theory of Change approach. The first approach must be able to accommodate those projects for which the Theory of Change is not very explicit, assumptions are not articulated, key variables for impact evaluation are not identified, and data on key indicators are not collected. In order to evaluate projects under these conditions, a Theory of Change approach should be represented by *logic models*.

For those GEF projects that do have well articulated interventions and expected results – and the assumptions that link the two, a more rigorous form of the Theory of Change approach should be used. This approach is best represented by *results chains*. Applying the more rigorous results chain approach allows evaluators to determine the extent to which projects have been successful. In particular it facilitates analysis of not only the extent to which a given intervention worked, but also why it worked or did not work.

The following figure is designed to assist the GEF EO and the evaluation team to decide when it is appropriate to use the logic model approach and when it is appropriate to use the results chain approach.



While attention should be paid to the activities and inputs side of theories of change in the course of conducting evaluations, to the extent possible, evaluators should always

strive to gauge impacts (Recommendation 8). If it proves impossible to assess impact – either through secondary or primary data – then evaluators should rely on outcomes data to determine the extent to which a project succeeded.

Clearly, there are many, many theories of change that could be created for a set of conservation projects. Given limited resources, however, the EO and evaluation team should work together with relevant GEF staff to determine the areas or themes of greatest interest (Recommendation 12).

Likewise, the evaluators should, to the extent possible, use a standard taxonomy of components of a Theory of Change approach (Recommendation 14). For a Theory of Change approach to work, evaluators must be able to quickly illustrate linkages between actions and outcomes/impacts in a way that is generally understandable to the conservation community and translatable across projects. To facilitate communication, it is therefore advantageous to have evaluators use a common nomenclature. This is true, in particular, for the main components of a Theory of Change approach: actions, threats/opportunities, and targets. The more unified and consistent the definitions of these components are across evaluations, the more comparable they will be. Evaluators should consider using the IUCN-CMP classifications found at the CMP website (www.conservationmeasures.org).

In addition to using the same terms and concepts, evaluators should use the same metrics to measure impact and changes in key factors along theories of change chains or logic models – especially to facilitate comparison of results across projects and the roll up of results to gauge higher levels of impact (Recommendation 15). While there are endless lists of indicators available from multiple sources, these lists tend to focus primarily on biodiversity targets and, to a lesser extent, on threats to biodiversity. In order to evaluate conservation interventions using a Theory of Change approach, it is important to analyze the path of association from intervention to impact. To do this, evaluators must have access to indicators related to variables that are found in the inputs, outputs, and outcomes portions of the chains as well. CMP is currently supporting work to develop the Strategic Indicator Selection System (StratISS), and the EO and evaluation team may wish to tap into this resource.

Data Collection

The pilot team should determine and develop questionnaires and data collection strategies that will permit them to collect the proper data to test the theories of change. The EO anticipates that the instruments will be primarily informal or key informant interviews. Data collection protocols will necessarily include compilation of secondary data as well as these direct interviews. Instruments should include strategies to evaluate direct impact, sociopolitical sustainability, programmatic sustainability, institutional sustainability, financial sustainability, and replication.

During the pilot phase, the evaluation team must work with GEF and IA staff to brief them on the underlying concepts inherent in a Theory of Change evaluation approach.

To do so, it is important to secure participation of key staff in the design and implementation of the pilot phase. GEF and IA staff should be invited to observe the development of the overall pilot approach and play a role in sample selection. To the extent possible, they should also participate directly in the pilot evaluations and accompany evaluators on field visits so that they can learn directly from the evaluation process about the projects and their impacts, as well as how to evaluate the projects.

It may be necessary for the evaluation team to consult with thematic experts to help develop appropriate logic models or results chains (Recommendation 13). Many of these experts will no doubt be available from the ranks of GEF and IA staff. It may be necessary, however, to consult with outside experts. Starting with expert-based theories of change will accomplish three things: It will help the EO base its work on the conventional wisdom in the field – starting with what is already known; it will make the process of determining which variables to use in the evaluation a more efficient process; and it will deflect any potential criticism of conducting evaluations not solidly founded in theory and knowledge.

Evaluation Matrix

Using the China protected area management strategy project as an example, a partial evaluation matrix might look like the following:

Figure 10. Illustrative Partial Evaluation Matrix for Protected Area Management Strategy (Conservation Education)

Key Questions	Indicators/Basic Data	Sources of Information	Methodology Components
1. Is conservation education improving the conservation status of giant pandas?			
- Is conservation education increasing knowledge?*	# or % community members knowledgeable about the importance of giant pandas	Conservation educators; community members	Review conservation educators' records; Survey community members
- Have community members' attitudes towards pandas become more favorable?*	# or % community members in favor of protecting pandas	Conservation educators; community members	Review conservation educators' records; Survey community members
- Are hunting incidences decreasing?	# fines to community members per month	Park guards' records	Review records
- Are pandas populations increasing/status improving?	# pandas Panda population structure Panda reproductive success	Local university ecology department	Transects – population counts (incl. by age and gender) Count of nesting sites

* Boxes missing from the results chain but ideally should be there, as they are important for laying out the assumptions

5. Pilot Workplan

Roles

There are three main roles or functions that will need to be carried out during the pilot: Conceptualization, Coordination, Management, and Oversight

These functions will be covered by EO team members. The EO will take a leadership role in providing the vision for the work and be responsible for identifying and contracting the necessary external assistance to carry out the pilot. In addition, the EO will identify the right people in GEF and the IAs that should participate in the pilot phase and any other related subsequent work.

While most of the technical work will be undertaken by consultants, the EO will provide management and oversight to maintain the quality of the work. This will ensure the pilot provides the feedback the GEF needs to adequately analyze the approach and improve it over time.

Collaboration, External Guidance, Liaison to Actual Projects

For the pilot to be successful, direct and significant involvement of key staff from the GEF biodiversity focal area and IAs is necessary. In particular, the EO anticipates that IA staff will help identify projects to be included in the pilot, set up field visits, help secure existing data, and act as a conduit between the evaluation team and the field.

GEF team members and IA technical staff will contribute on the technical side as well. For example, these team members may help create the initial theories of change, comment on the development of data collection instruments and protocols, and provide insights into what data to collect.

Development of Theories of Change and Instruments, Collection and Analysis of Data, and Write-up

The consultant, in coordination with the EO, will be responsible for researching and developing chains, developing instruments, conducting informal interviews with key informants, and carrying out field visits to collect data and information. In addition, the consultant will be responsible for initial drafts of reports and refinement of those reports based on feedback from the EO, other GEF staff, and IA representatives.

Steps

The EO foresees eight main steps in the implementation of the pilot phase of the Theory of Change approach (see Attachment 2 for an illustrative timeline):

Step 1. Assemble the pilot team

The EO will take the lead on identifying appropriate team members to participate in the pilot of the proposed evaluation approach. The EO anticipates including three GEF projects in the pilot phase. Each pilot site will be selected from the biodiversity focal area portfolio and associated with at least one IA. Projects will be selected in a way that maximizes potential thematic overlap in order to test the assumption that it will be possible to aggregate results from various evaluations, across different sites. In addition, the EO will need some external help with certain aspects of the pilot as it will be very time intensive. As such, we recommend that the EO compose a pilot team of EO, biodiversity focal area, and appropriate IA staff and consultants to carry out the work.

The EO will coordinate and manage the process, biodiversity focal area; IA staff will primarily advise; and the consultants will do the bulk of the technical work required to complete the pilot phase.

Step 2. Refine the pilot protocol

The pilot team will take the results of the GEF Impact Evaluation study completed in September 2006 and refine them, as needed, to operationalize the pilot. The pilot team should discuss and agree on the final selection of theme and projects to be included in the pilot and the general approach developed in this report. Based on its analysis, the pilot team should refine and adapt the approach as it sees fit.

Step 3. Draft theories of change and candidate indicators

After the final list of projects to be included is completed, the pilot team members should develop draft theories of change and the indicators required to test them. Special attention should be paid to develop theories of change and indicators that are consistent across projects in order to test where data from various sites can be combined during the analysis phase. For example, if the pilot focuses on protected areas management strategies, it would be useful to develop theories of change related to PA infrastructure development, PA patrols and vigilance, and alternative development strategies for buffer zone communities.

The pilot team should review project documents and other secondary data sources. They should also interview key individuals, including past project managers and IA/biodiversity focal areas staff who may have worked on the project. From these sources of information, the pilot team should determine on which specific actions – within the chosen theme – they will concentrate for the pilot (e.g., PA infrastructure, environmental awareness campaigns for neighboring communities). Once this has been accomplished, the team should draft the theories of change that they will use as the foundations of their evaluations of the pilot projects.

Step 4. Prepare data collection instruments

Once the team has determined the specific themes and strategies it will address and the projects it will include, it should develop data collection instruments. The EO anticipates the most useful methods will be informal/key informant interviews, direct observation, and some group interviews. Given budgetary limits and time constraints for the pilot, other, more time intensive and costly methods will not be appropriate. The data collection instruments should include questions designed to address the key questions outlined in the Scope section related to direct impact, sociopolitical sustainability, programmatic sustainability, institutional sustainability, financial sustainability, and replication.

Step 5. Interview key informants

The pilot team should ground truth and refine the draft theories of change and instruments with key individuals in the biodiversity focal area and respective IAs. Team members should also contact key individuals in the field that may provide information that would lead to further refinement before arriving on-site.

Step 6. Carry out field visits

By the time the team representatives are ready to go the field, they should have well-grounded instruments and theories of change and a clear method designed to make data collection as efficient as possible. Every effort should be made to contact and interview key individuals who are still at the project site (e.g., PA managers, partner organizations working in the area, community leaders). The evaluators should also conduct some direct

observation to further triangulate results. The EO estimates that two weeks in the field (project area) plus two weeks in national and regional centers will be required per project.

Step 7. Analyze data and information and present results

After the completion of field visits and other key informant interviews as needed, the team should analyze the data and information that were collected. The consultants will be primarily responsible for this, but they should work very closely with the entire pilot team. It will be important to share the results with the rest of the EO and key members of the biodiversity focal area and IAs in order to highlight those issues that are most important and relevant to GEF.

The pilot team should also focus on aggregating the results of the piloting exercise across projects. One purpose of the pilot is to see if theories of change and indicators cut across multiple sites – and to see if it is possible to aggregate results across sites to learn more efficiently about what works under which conditions. So, for example, if the GEF is supporting 50 protected areas, what can it say about its impact on reducing threats to the PAs or protecting the biodiversity at those sites? If 20 of those projects use environmental education, what can the GEF say about the overall success of that strategy and under conditions that strategy tends to work well or not work well? This will be a crucial element in the analysis phase of the pilot.

Step 8. Prepare final report

Based on feedback from the IAs and biodiversity focal area staff, the pilot team should prepare a final report outlining what it learned about the piloting of the Theory of Change approach, including recommendations for moving forward.

6. Key Audience and Partners Involved

The initial audience targeted is the GEF Council, which has a specific interest in the issue of impacts. However, it appears likely that there are differing expectations and motivations among Council members in pressing for impact evaluations. The pilot study (together with a summary of the WB GEF impact evaluations) should therefore be used to draw out the key perspectives of Council members, to allow the main phase of the studies to deliver products, which are likely to satisfy the major expressed needs of Council.

GEFSEC and the IAs also have a major interest in establishing the impacts of the GEF's work. The initiative of the World Bank GEF Office in this respect provides a valuable body of work, which will be drawn upon by the EO study. Discussions will be held during the pilot phase to establish the extent of interest of the IA's Evaluation Offices in collaborating with the GEF EO pilot and main studies.

The broader environmental community, including the major international NGOs, is also likely to have a strong interest in the work to be undertaken and one possibility would be to host or co-host, as an integral part of the pilot study, a workshop to share the experience of impact evaluation of environmental interventions within this broader community. With regard to biodiversity, for example, this could be linked to the process to develop indicators under the CBD. This has not been budgeted.

Country level partners, particularly in countries receiving large GEF inputs, should also have considerable interest in assessing the impacts of the GEF portfolio, particularly in view of the results-based intentions of the RAF.

The impact results would be an important input into OPS4, enabling this to say something about long term results of the GEF's work for the first time.

Attachment 1

Sample of Projects Identified for Inclusion in the Analysis

Pre-GEF-3 Projects*

20 Completed Projects							
GEF_ID	IA	CountryName	ProjectName	Type	Phase	Region	Included?
136	World Bank	Ghana	Natural Resource Management	FP	GEF - 1	AFR	Yes
78	World Bank	Lao PDR	Wildlife and Protected Areas Conservation	FP	Pilot Phase	Asia	Yes
90	World Bank	Russian Federation	Biodiversity Conservation	FP	GEF - 1	ECA	No
62	World Bank	Mexico	Protected Areas Program	FP	Pilot Phase	LAC	Yes
220	UNDP	Comoros	Conservation of Biodiversity and Sustainable Development in the Federal Islamic Republic of Comoros	FP	GEF - 1	AFR	Yes
209	UNDP	Vietnam	Vietnam PARC - Creating Protected Areas for Resources Conservation (PARC) in Vietnam Using a Landscape Ecology Approach	FP	GEF - 1	Asia	Yes
197	UNDP	Guatemala	Integrated Biodiversity Protection in the Sarstun-Motagua Region	FP	GEF - 1	LAC	Yes
173	UNEP	Global	Global Biodiversity Assessment	FP	Pilot Phase	CEX	No
142	UNEP	Global (Brazil, China, Ghana, Guinea, Kenya, Papua New Guinea, Tanzania,	People, Land Management, and Environmental Change (PLEC)	FP	GEF - 1	CEX	No
125	UNDP/World Bank	Madagascar	Environment Program Support Project	FP	GEF - 1	AFR	No
83	World Bank/UNDP	China	Nature Reserves Management	FP	GEF - 1	Asia	Yes
1242	UNEP/UNDP	Regional (Burkina Faso, Botswana, Kenya, Mali, Namibia, Niger, Senegal,	Desert Margin Programme, Phase 1	FP	GEF - 2	AFR	No
539	World Bank	Poland	Forest Biodiversity Protection	FP	Pilot Phase	ECA	No
592	UNDP	Belize	Conservation And Sustainable Use of the Barrier Reef Complex	FP	GEF - 2	LAC	Yes
216	UNDP	Lebanon	Strengthening of National Capacity and Grassroots In-Situ Conservation for Sustainable Biodiversity Protection	FP	GEF - 1	Asia	Yes
206	UNDP	Uruguay	Consolidation of the Banados del Este Biosphere Reserve	FP	GEF - 1	LAC	No
54	World Bank	Uganda	Bwindi Impenetrable National Park and Mgahinga Gorilla National Park Conservation	FP	Pilot Phase	AFR	No
541	UNDP	Regional (Kenya, Tanzania, Uganda)	Reducing Biodiversity Loss at Cross-Border Sites in East Africa	FP	GEF - 1	AFR	No
79	World Bank	Philippines	Conservation of Priority Protected Areas	FP	Pilot Phase	Asia	No
538	World Bank	Peru	National Trust Fund for Protected Areas	FP	Pilot Phase	LAC	Yes

* Those projects indicated in the “Included” column were included in the final analysis.

GEF-3 Projects

10 Projects Approved in GEF3

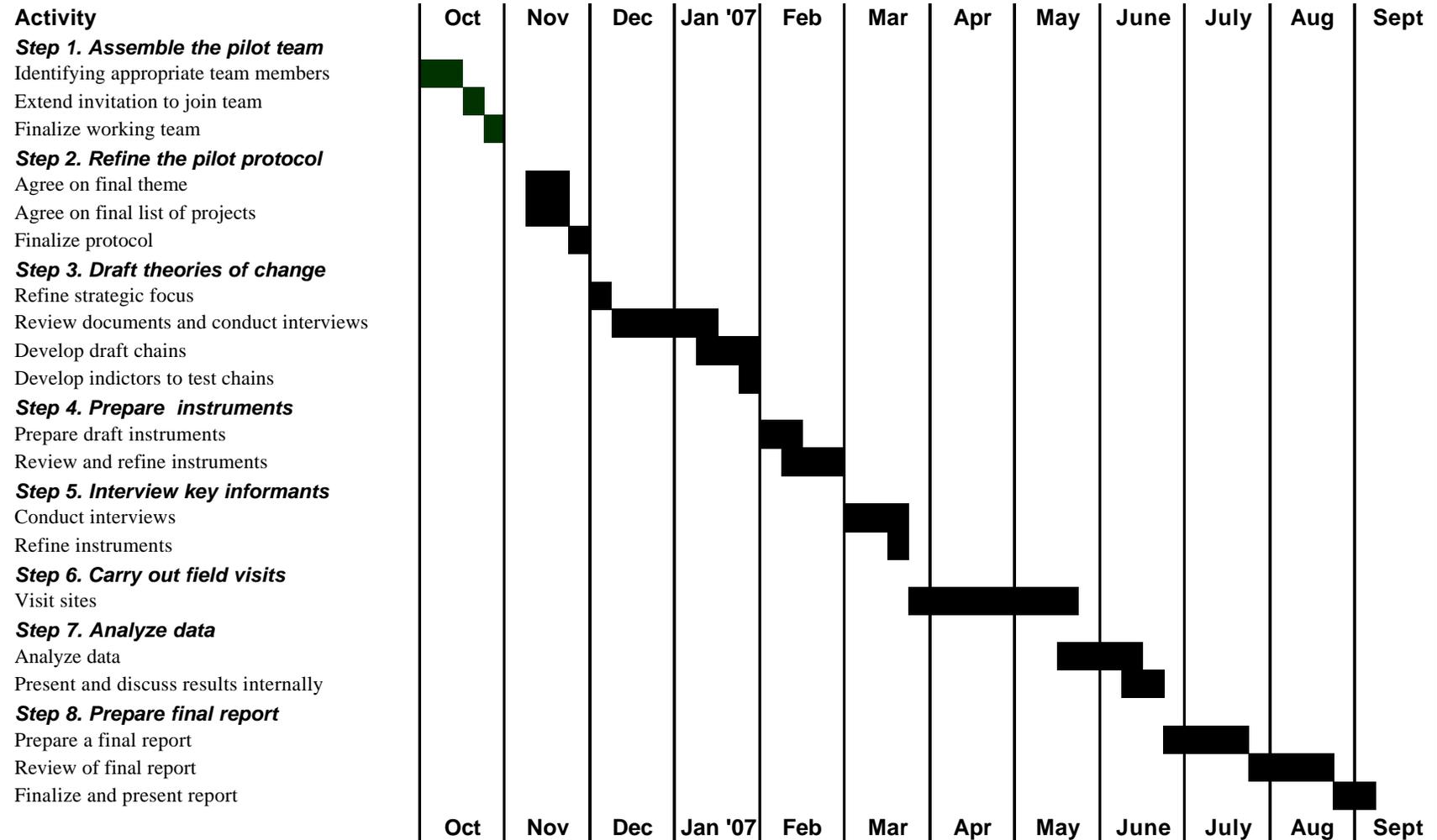
GEF_ID	Agency	Country	Project Title	region
2764	World Bank	Brazil	National Biodiversity Mainstreaming and Institutional Consolidation Project	LAC
2101	World Bank	Tanzania	Marine and Coastal Environment Management Project (MACEMP)	AFR
1156	UNDP	India	Mainstreaming Conservation and Sustainable Use of Medicinal Plant Diversity in Three Indian States	Asia
2104	UNDP	Belarus	Catalyzing Sustainability of the Wetland Protected Areas System in Belarusian Polesie through Increased Management Efficiency and Realigned Land Use Practices	ECA
1097	UNEP	Regional (China, Iran, Kazakhstan, Russian Federation)	Development of a Wetland Site and Flyway Network for Conservation of the Siberian Crane and Other Migratory Waterbirds in Asia	REG
1918	UNEP	Regional (Colombia, Ecuador, Venezuela, Peru)	Conservation of the Biodiversity of the Paramo in the Northern and Central Andes	LAC
1126	ADB	China	Sanjiang Plain Wetlands Protection Project	Asia
1515	IADB	Honduras	Consolidation of Ecosystem Management and Biodiversity Conservation of the Bay Islands	LAC
1184	World Bank	Jordan	Conservation of Medicinal and Herbal Plants	Asia
1053	UNDP	Regional (Gambia, Guinea, Mali, Senegal)	In-situ Conservation of Endemic Ruminant Livestock in West Africa	AFR

All Projects included in this table were included in the analysis.

Attachment 2

Illustrative Timeline

October 2006 – September 2007



Attachment 3

Illustrative Budget

Item	Unit	\$/Unit	Number	Amount
Consultant time				70,000
<i>Develop Approach</i>	<i>per day</i>	<i>1,000</i>	<i>25</i>	<i>25,000</i>
<i>Write Up Final Approach</i>	<i>per day</i>	<i>1,000</i>	<i>25</i>	<i>25,000</i>
<i>Work w/ Field Testing Office</i>	<i>per day</i>	<i>1,000</i>	<i>20</i>	<i>20,000</i>
Field Test				
<i>(3 trips of 5 days in country to Asia, Africa, Latin America for a team of 2)</i>				
LAC				11,500
<i>Airfare</i>	<i>avg per trip</i>	<i>3,000</i>	<i>2</i>	<i>6,000</i>
<i>Local transportation</i>	<i>avg per day</i>	<i>250</i>	<i>10</i>	<i>2,500</i>
<i>Per Diem</i>	<i>avg per day</i>	<i>250</i>	<i>10</i>	<i>2,500</i>
<i>Miscellaneous travel expenses</i>	<i>avg per person</i>	<i>250</i>	<i>2</i>	<i>500</i>
Asia				13,500
<i>Airfare</i>	<i>avg per trip</i>	<i>4,000</i>	<i>2</i>	<i>8,000</i>
<i>Local transportation</i>	<i>avg per day</i>	<i>250</i>	<i>10</i>	<i>2,500</i>
<i>Per Diem</i>	<i>avg per day</i>	<i>250</i>	<i>10</i>	<i>2,500</i>
<i>Miscellaneous travel expenses</i>	<i>avg per person</i>	<i>250</i>	<i>2</i>	<i>500</i>
Africa				11,500
<i>Airfare</i>	<i>avg per trip</i>	<i>3,000</i>	<i>2</i>	<i>6,000</i>
<i>Local transportation</i>	<i>avg per day</i>	<i>250</i>	<i>10</i>	<i>2,500</i>
<i>Per Diem</i>	<i>avg per day</i>	<i>250</i>	<i>10</i>	<i>2,500</i>
<i>Miscellaneous travel expenses</i>	<i>avg per person</i>	<i>250</i>	<i>2</i>	<i>500</i>
Communication				1,000
Materials / Supplies				500
TOTAL				108,000