The Wildlife Conservation Project: A Literature Review*

Jonathan Mawdsley
Robin O'Malley
The Heinz Center

Introduction

We have written this review in an attempt to summarize the "state of the art" in developing performance measures for wildlife conservation activities, drawing on a substantial body of both peer-reviewed and "gray" literature, as well as interviews and discussions with numerous conservation practitioners and evaluation professionals (listed in the "Acknowledgments" section below).

The focus of this review is not on "how to measure," but rather on "how to develop measures." As most wildlife biologists are aware, there is a substantial body of primary and secondary literature that describes best available practices and statistically valid approaches for measuring or monitoring wildlife populations, habitats, and ecosystems. The real challenge for practitioners and managers is that there are a very large number of environmental attributes that could potentially be measured in a rigorous evaluation of any one particular management action. Resources for monitoring and evaluation are usually quite limited, which means that only a few things can be measured well enough to provide data at the level of accuracy needed to inform management decisions. This means that managers must choose a few indicators to measure, and choose them with considerable care.

In the sections that follow, we discuss methods and tools that evaluation professionals have used to guide practitioners through the process of selecting appropriate indicators and metrics for project monitoring and evaluation. We also describe the general types of environmental attributes that are most commonly selected in monitoring and evaluation schemes for wildlife conservation projects. A list of sample indicators, derived from the current literature and from our conversations with practitioners, is provided in the Appendix. Finally, we provide an annotated list of references for those interested in exploring particular topics in greater depth.

*** With some original thoughts thrown in for good measure.**

1.) Why measure?

We live in a performance-driven society, where individuals, businesses, governments, and social service organizations are all routinely subjected to performance assessment and evaluation. The practice of managing wildlife is certainly not exempt from this broader trend. Indeed, state and federal legislators, government agencies, and private funders are increasingly asking wildlife professionals to explain how – or whether - their management activities have led to measurable improvements in animal populations.

As any wildlife manager knows, answering even the most basic questions about the results of management activities can be surprisingly difficult (Walter 1986; The Wildlife Society 2002). Most conservation activities are not designed as controlled experiments to rigorously establish cause-and-effect relationships between management actions and wildlife responses. Without a rigorous experimental design, a manager is confronted with seemingly endless confounding factors that make it hard to say for certain whether observed changes in wildlife populations result from any particular management action. One of the most significant confounding factors is the time lag that often occurs between management activities and wildlife population responses. This time lag obviously makes it difficult to measure results on annual or quarterly time scales, as may be required by government or agency reporting programs. To further complicate matters, standard monitoring programs for rare species or game species are often not set up to provide data at a frequency or spatial scale necessary for measuring the performance of specific management actions.

Wildlife managers are not the first group of professionals to encounter these kinds of difficulties. Similar problems are shared by practitioners in social service, business, and health care sectors (Fazey et al. 2004; Pullin and Knight 2001; Stem et al. 2005; Trochim 2006). These fields share certain commonalities with wildlife management: the complexities of the populations with which they work; a desire to commit resources towards “actions” rather than towards monitoring and results; the difficulty of linking very specific, small-scale actions to the desired “big picture” results; and the challenge of rigorously evaluating project and program outcomes on limited monitoring budgets.

In response to these common problems, a new science of evaluation has developed in recent years (Fazey et al. 2004; Kleiman et al. 2000; Salafsky et al. 2002; Stem et al. 2005; Trochim 2006). This report has been written to help wildlife managers become more conversant in the language and methods of this new science. There is much to recommend this new perspective: managers can use evaluation techniques to document and demonstrate their successes, and to understand and learn from projects that do not succeed. Adopting an evaluation perspective means that wildlife professionals may need to think in new ways about conservation activities: clearly defining the logical steps between activities and big-picture goals, linking each of these steps to potential indicators and metrics, and selecting a small suite of metrics (from among the many that are possible for each project) that most effectively and efficiently measure accomplishments.

Another important element of this new thinking involves a careful consideration of the existing programs that monitor wildlife populations. Some data from these programs are useful or relevant to performance measurement, while other data are not.

Although monitoring has long been recognized as an important aspect of wildlife management, many of the existing monitoring programs have not been set up in a way that will enable managers to test whether their activities are related to changes observed in wildlife populations.

A question that has come up repeatedly in our conversations with wildlife managers is whether or not this recent focus on evaluation is simply a fad, or whether it is going to become an established part of doing business. While it is hard to predict the future, there are a number of signs that suggest that evaluation and performance measurement are here to stay. The Government Performance and Results Act of 1993 (GPRA) requires all federal agencies to track their performance, and report their progress towards achieving meaningful goals. The Program Assessment and Rating Tool (also known as PART), originally developed as part of the implementation of GPRA, is being applied by the White House Office of Management and Budget to all federal programs dealing with wildlife, including programs under the U. S. Fish and Wildlife Service, NOAA Fisheries, USDA Forest Service, USDA Natural Resources Conservation Service, the Bureau of Land Management, and the National Park Service. Many state legislatures have also passed similar accountability legislation that requires regular progress reports from state agencies, including those responsible for wildlife management. And in the private sector, many foundations and other private funders are implementing their own evaluation protocols which require groups that receive funding to track their own accomplishments.

Wildlife managers who are interested in implementing some form of evaluation or effectiveness management can now choose from a number of potential tools and systems for conducting evaluations (Stem et al. 2005). The second section of this review describes some of the more common approaches, from logic frameworks and causal chains to adaptive management cycles. From conversations with practitioners, it is clear that there is no right tool or approach for an individual project; rather, what matters is that one has a clear understanding of the scientific rationale for the project and collects data at appropriate times and geographic scales to be able to document changes resulting from project implementation.

One of the common features of all of the systems that have been proposed for evaluation or performance measurement is the use of indicators – simple statistics that are measured at intervals and can be used to track the project's results. The third section of this review discusses indicators and metrics (specific things or ways of measuring an indicator) that are commonly used to assess the results of species and habitat conservation projects, respectively. One of the interesting findings of this report is that there are a relatively modest number of "things" that are commonly measured by practitioners.

Finally, while it is good to be able to measure the effectiveness of individual conservation actions, it is also important for managers to be able to relate the outcomes of their small-scale actions to broader environmental trends that are being observed in the larger world. Conservation partners and donors may care about individual project-specific outcomes – whether there are more savannah sparrows at a site, for example, but they also care about whether we are contributing to major societal goals such as "saving biodiversity" and "preventing species extinction." The fourth section of this review examines challenges and opportunities associated with linking project- and program-specific metrics to some of the available national environmental indicators, such as those contained in The Heinz Center's "State of the Nation's Ecosystems" report (Heinz Center 2002) or the U. S. Environmental Protection Agency's "Report on the Environment" (U. S. Environmental Protection Agency 2003).

In summary, then, evaluation and performance measurement are valuable tools for wildlife managers. They provide frameworks and structures for telling scientifically credible stories about the anticipated outcomes of conservation actions. They help identify strategies that work, and can also suggest possible improvements to strategies that don't quite work as planned. Wildlife managers know, intuitively, that they have done good work for wildlife and been good stewards of the resources that society has dedicated to wildlife conservation. Evaluation and performance measurement can help managers make a clear and logical argument that this is indeed the case.

2.) How do you develop measures?

As mentioned in the introduction, there are many different attributes of wildlife populations, ecosystems, and habitats that could potentially be measured to determine whether or not management activities are having the desired effect. In this section, we examine five different approaches for choosing among the universe of possible indicators and metrics. Each approach starts from a different place: goals and objectives, audience information needs, activities, outcomes, or threats.

We would like to emphasize that the development of performance measures is just one aspect of a more comprehensive process of project planning, implementation, monitoring, and evaluation. For more comprehensive information about other aspects of project development and implementation, the reader will find it helpful to consult references such as Margoluis and Salafsky (1998), Walters (1986), Salafsky et al. (2002), and The Wildlife Society (2002).

In addition to the general approaches discussed below, Hagan and Whitman (2004) review indicators that have been developed for a wide spectrum of sustainable forestry projects and conclude that successful indicators need to meet all five of the following criteria. Indicators should be **relevant** (changes in the indicator reflect changes in the underlying ecological process or state that we are trying to measure), have **scientific merit** (be supported by best current science), have **ecological breadth** (encompass a number of species, structures, and/or processes), be **practical** (feasible to measure given available resources of time, money, and skill), and have **utility** (able to inform decision-makers about needed actions). These authors recommend that indicator selection begin with the clear identification of the project's goals, objectives, and spatial scale, and then proceed through a clear, transparent, and inclusive process that involves scientists, forest stakeholders, and forest managers in the actual process of indicator selection. Potential indicators that are identified through this process and meet all five of the criteria above should be presented to managers with the question of how they might actually make decisions using the information from the indicator (if the answer is unclear, the indicator is likely of little utility and should be discarded). These authors also recommend that the final indicator set be kept as small as possible, and recommend looking for single indicators that cover multiple biodiversity values.

2. A.) Goal, Objective and Audience-driven Approach

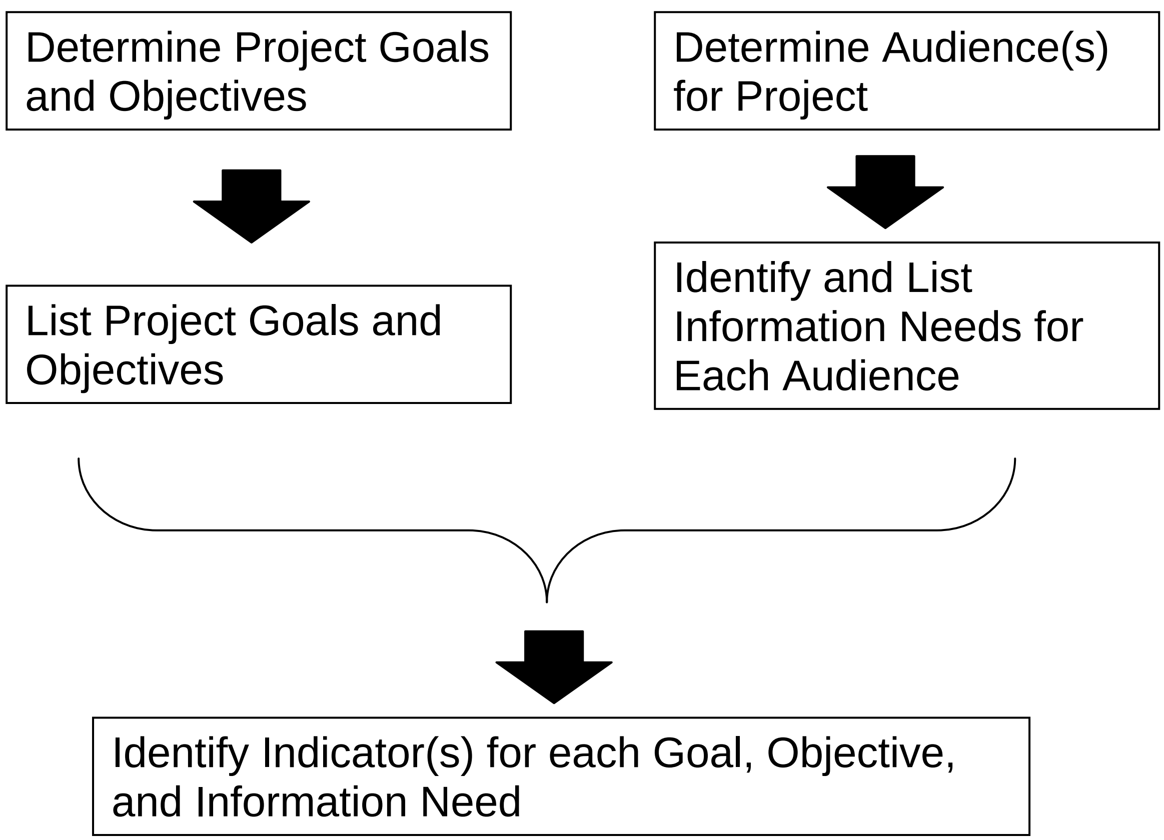
In their excellent book "Measures of Success," authors Richard Margoluis and Nick Salafsky (1998) take readers through the process of designing and implementing monitoring and evaluation plans for biodiversity conservation projects. These authors recommend linking performance measures to the project's **goals and objectives**, as well as the **information needs of the project's audience.**

Following Margoluis and Salafsky (1998), **goals** are "a general summary of the desired state that a project is working to achieve" (e.g. "Conserve biodiversity in the Kalimantan rainforest" or "Reduce nutrient inputs to Chesapeake Bay") while **objectives** are "specific statements detailing the desired accomplishments or outcomes of a project" (e.g. "Reduce illegal rainforest logging by 50% over 5 years," "Implement nutrient reduction projects on half of the farms in the Shenandoah Valley over the next 10 years”). Each goal may have one or more objective associated with it.

Margoluis and Salafsky (1998) further recommend that project managers identify key audiences for a project (both internal audiences and external audiences) and list key information needs for each audience. Information needs may be very specific (the number of whales in a bay), or broad (an overall sense of whether or not a project is worth the money expended on it).

Once goals, objectives, and key information needs for the project’s major audiences have been identified, **indicators** are selected for each of these items. In this system, indicators are environmental or social attributes that can be measured and that change over time during the course of a project or program.

The following flowchart summarizes the method for choosing indicators described by Margoluis and Salafsky (1998):



These authors recommend that indicators be **measurable, precise, consistent,** and **sensitive**. Measurable means that an indicator can be reported and analyzed using either qualitative or quantitative methods. Note that resource limitations may also dictate whether or not a particular indicator is measurable! Precise means that there is general agreement among practitioners as to how a particular indicator is defined. Consistent means that the indicator does not change over time; it always measures the same thing. And sensitive means that changes in the indicator are proportionate to changes in the environmental condition or item being measured: a large change in the indicator reflects a large change in the environment, while a small change in the indicator reflects a small change in the environment (Margoluis and Salafsky 1998).

2. B.) Activity-based Approach: Logic Framework

The logic framework is a commonly used tool for performance measurement and project assessment. It differs from the method recommended by Margoluis and Salafsky in that it links performance measures to specific project activities, which are in turn linked to the project's goals and objectives. Unlike the Margoluis and Salafsky method, the information needs and audiences for the project are not explicitly considered.

The logic framework was first developed in the 1970s by the U. S. Agency for International Development, as a tool to help grant applicants explain their projects more clearly. In its most basic form, the framework asks project managers to list their project's goal and specific objective, describe specific activities that they will be undertaking to attempt to achieve an objective, and predict the specific results that they expect will follow from those activities. Some versions also ask a project manager to explain how s/he would measure the results of the activities s/he is proposing.

The logic framework is useful for showing simple relationships between actions and anticipated results. Completing even a simple logic framework for a project can help managers tell a more compelling story about the work that they are doing, and describe both the short-term and long-term results that they are expecting to achieve.

Many funding agencies and private foundations in the United States are using logic frameworks and are asking their grantees and applicants to use this tool in developing their proposals. The following example is adapted from materials that have been developed by Matthew Birnbaum at the National Fish and Wildlife Foundation, a private non-profit organization that provides significant financial support for wildlife conservation activities in the United States. We have used this example because it is likely to be familiar to many wildlife professionals who seek funding from this foundation.

Here, the logic framework takes the form of a simple table or chart with seven columns and row(s) for each of the project's major activities. In this example, there are no columns on the left-hand side for "Goals" or "Objectives" because the grant application form asks for the project's goals and objectives in a different section. For the sake of completeness I have listed these two essential components above the logic framework table.

Project: Demonstration Planting, Forested Riparian Buffer
Goal: To reduce non-point source water pollution from agricultural sources in Adams County, Pennsylvania.
Objective: To stabilize 50 feet of eroding streambank and plant a new riparian forest buffer along Rock Creek adjacent to Farmer Jones' cow pasture.

Activity	Output	Outcome	Indicator	Baseline Value	Output Value	Outcome Value
Stabilize eroding streambank	Streambank stabilized	Erosion stopped	Number of feet of eroding streambank	50	0	0
Plant 50' Riparian Buffer	100 new trees planted	Bank stabilized	Number of trees on bank	0	100 at project completion	75 trees surviving after 5 years

This logic framework actually has two parts or sections – the three columns on the left-hand side, which provide a short narrative description of the project's activities and results, and the four columns to the right, which describe how the results of the project will be measured using a simple statistic or "indicator."

To make this example clearer, here are some formal definitions for the terms at the top of each of the columns.

Activities – Actions that will be undertaken as part of a project or program. The more specific, the better.

The next two columns – outputs and outcomes – are results that should follow logically because of the activity/ies that are listed in the first column.

Output – These are short-term results that should happen immediately or very soon after the activities in the first column are implemented as planned. Outputs are often quantitative (number of farmers contacted, number of acres treated for weeds, number of miles of river opened to fish passage) although it's possible to have non-quantitative outputs (for example, dam removed).

Outcome – Longer-term, "big picture" quantitative or qualitative results that are anticipated if the activities in the first column are implemented as planned. Outcomes should be related to the larger goal for the project.

Indicator – In this context, an indicator is an aspect of the environment that can be measured to determine whether or not the project has met its objectives. Note that there is a difference between indicators and **metrics** (a metric is a specific way of measuring a particular environmental attribute). See Margoluis and Salafsky (1998) for additional discussion.

Baseline, output, and outcome values – These three columns list the actual or anticipated measurements of the project's indicator at three distinct times: when the project is being proposed (baseline), after the project has been implemented (output) and after all benefits have been achieved (outcome).

Logic frameworks are useful in developing simple stories about a project and what would be expected to happen as a result. A simple story using the information contained in the logic framework above would look something like this:

"The Friends of Rock Creek are interested in reducing non-point source water pollution from agricultural sources in Adams County, Pennsylvania. To help meet this goal, we will stabilize 50 feet of eroding streambank and plant a new riparian forest buffer adjacent to Farmer Jones' pasture. These activities will stop erosion at this site.

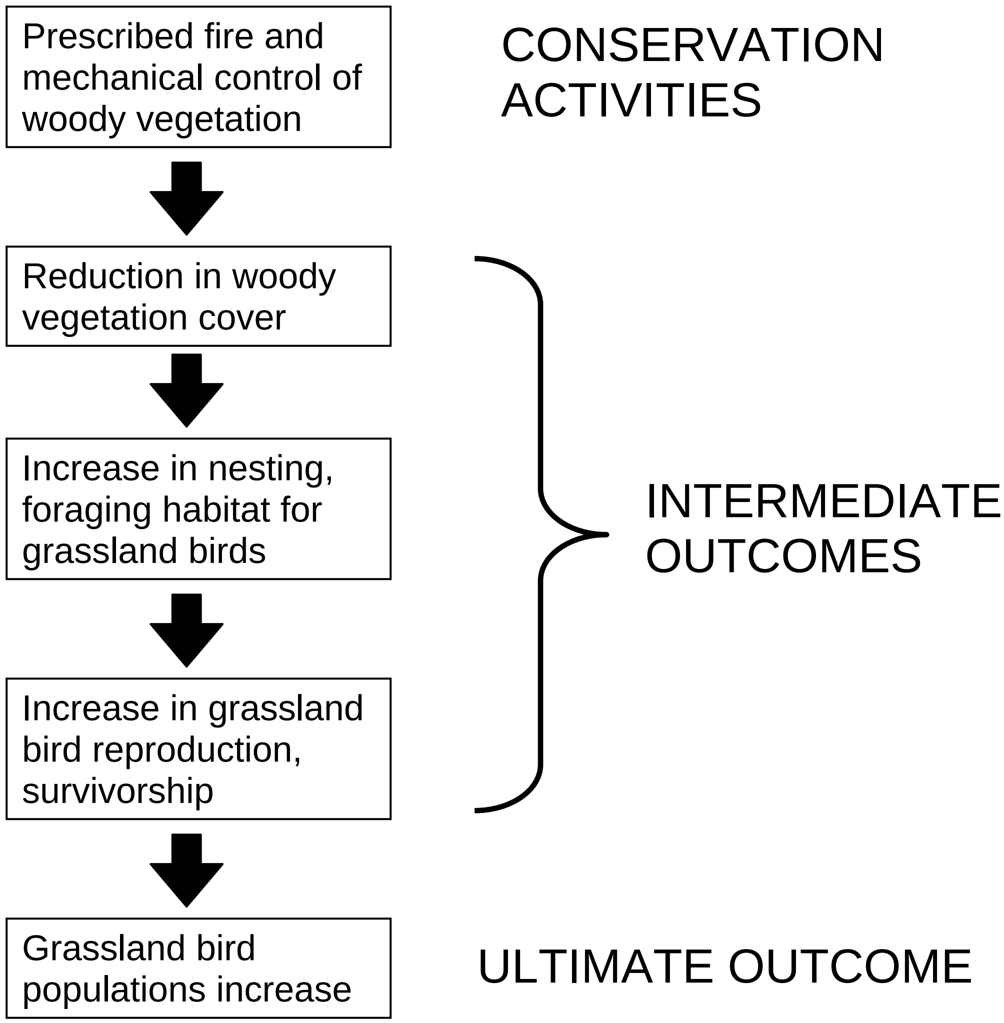
The erodec streambank currently has no trees; by the end of the project we will have planting 100 new trees at this site. We expect 75% of the trees to survive at least 5 years."

Notice that this story tells what the group is planning to do, what will happen as a result in both short- and long-term time scales, how they will measure these results, and even what they expect the measurements will be. After the project is completed, the project manager can compare her estimates with real-world data (collected through a monitoring program) to see how well her expectations aligned with the actual observed results.

2. C.) Activity and Outcome-based Approach: Causal Chain

Another useful tool is the causal chain (also known as a results chain) which expands on the simple logic framework by adding the full sequence of logical steps between an activity and the big-picture or long-term outcome for the project (Margoluis and Salafsky 1998). Causal chains start by listing a specific action or activity at the top of a piece of paper. At the bottom of the piece of paper is the project's goal. Between the activity and the goal are listed as many intermediate steps as needed to link the two in an unbroken logical progression. In completing the chain, it is helpful to keep asking the question "and then what happens" at each step until the activity and goal are completely linked in a chain of logical steps.

Here is an example of a causal chain for a project intended to increase grassland bird populations:

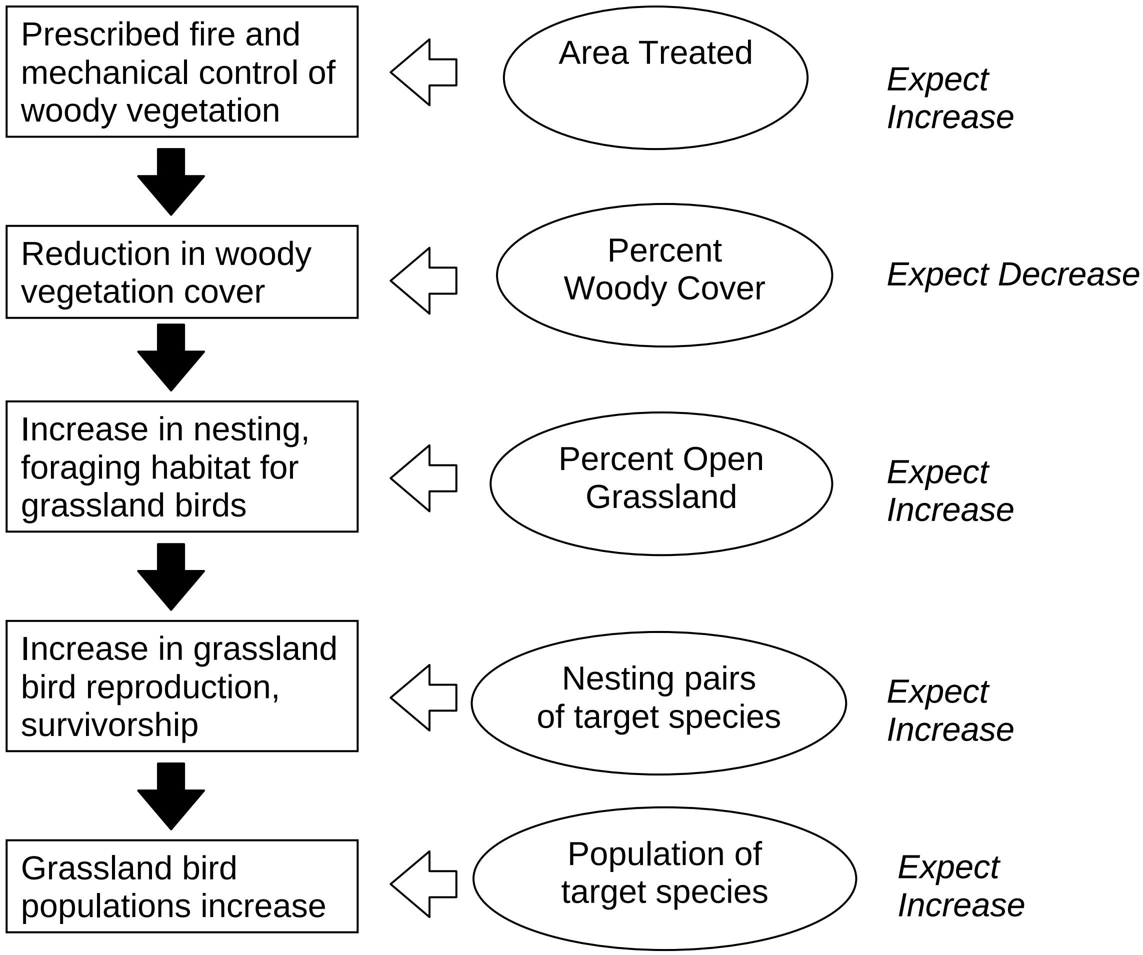


Here there are three steps between the specific conservation activities that the manager is planning to undertake, and her big-picture goal for doing these activities in the first place. The activity statement, goal statement, and intermediate steps could be made even more detailed and specific to fit a particular project.

Causal chains are even more useful than logic frameworks in developing a clear and compelling story about a project. Here's an example of a story that could be developed from the causal chain above:

"The goal of this project is to increase grassland bird populations at our prairie preserve. We will implement a vegetation management regime that includes prescribed fire and mechanical treatments to reduce woody vegetation. As a result of these treatments, we expect increases in nesting and foraging habitat for grassland birds, which should lead directly to increased survivorship and nesting success. We expect that these factors will contribute to an increase in the population of these bird species at our preserve."

A well-developed causal chain can also be of considerable assistance in designing a monitoring program for a project. Here is the grassland bird causal chain again, but this time with a list of potential indicators (in ovals) that could be measured by a project manager to determine whether or not the project had the desired effect. To the right of these indicators, our project manager has listed the trends that would be expected in each of her indicators if the project was implemented.



Note that even though the manager has listed potential indicators, she is still one step removed from selecting a "metric" or specific environmental attribute that will actually be measured in a monitoring program. This is because there are often multiple ways to measure a particular indicator. For example, "percent woody cover" or "percent open grassland" could be estimated using digitized aerial photography, or extrapolated from measurements made on the ground using a series of sampling plots. The population size and number of nesting bird pairs could be estimated using data from sample plots, transect walks, or determined directly from a complete census (which is usually only feasible for small sites).

This causal chain also shows that some indicators are closely related and could probably be combined in an actual monitoring scheme. For example, percent woody cover and percent open grassland are complementary for many grassland sites, meaning that an increase in one of these indicators is accompanied by a decrease in the other, and vice versa. Likewise, the number of nesting pairs of a bird species may be closely related to the overall population size, and may be easier to determine than overall population size for certain species in which males are brightly colored and/or exhibit elaborate courtship displays.

Developing a causal chain for a project can also help in making a choice among multiple possible indicators. In real-world situations, budgetary constraints often limit the size of monitoring programs, meaning that only a few of the numerous potential indicators (and even more numerous metrics) can be actually measured. The causal chain shows which of the many possible indicators are closest in logical proximity to the goal for the project. If a manager can only measure one thing, it stands to reason that she would want to measure something that directly reflects whether or not a goal has been achieved. For the case above, this would mean focusing monitoring resources on measuring the bird populations.

However, it may not always be possible to measure a project's contribution to the primary goal directly. While conservation projects are often focused on particular wildlife populations, it may be difficult or expensive to obtain accurate population estimates let alone the trend data that would provide longer-term feedback on the success of management activities. There may also be significant time lags between when an action is taken and when a response is seen in the wildlife population of interest. By backing a step or two up the causal chain, it may be possible to identify a "proxy indicator" associated with an earlier step in the chain that nonetheless provides valuable information about whether or not a project is achieving (or is likely to achieve) its full environmental benefits.

Dam removal projects provide a good example of a proxy indicator. We know that removing a dam opens up a stretch of river to migratory fish species whose movements were previously blocked or impeded by the dam. Yet counting the actual numbers of fish moving through a restored river is expensive and requires specialized equipment and expertise that is not widely available to local conservation groups. In this case, the number of stream miles opened to fish passage is a widely used proxy measure for the number of fish.

2. D.) Threat-based Approach

Salzer and Salafsky (2006) describe an interesting new conceptual approach for indicator selection that focuses primarily on the threats facing a particular natural area. These authors recognize two different types of indicators: "early warning" indicators that provide advance warning of potential problems, and "diagnostic indicators" that provide information on whether or not a specific action is working as planned. It should be noted that this is primarily a functional distinction rather than a substantive distinction, and that similar or even identical aspects of the environment may be measured as either "early warning" or "diagnostic" indicators.

The Salzer and Salafsky method can be applied at any scale but assumes that a specific area for management has been defined. The first question confronting managers is whether specific and substantial threats are facing this area, necessitating management action. If there are no threats, the next question is whether there are known potential threats. If no, then only early warning indicators would need to be measured. If there are known potential threats, then both early warning indicators and diagnostic indicators specific to those threats should be measured.

If however there are substantial threats facing the area of interest, then the next question is whether or not there are clear and feasible actions to abate these threats. If the answer is no or even "probably not," then the only course of action would be to apply diagnostic indicators to measure the status and progress of the threat. If there are actions that *might potentially* abate the threats, then the best course of action would be to implement a small-scale test to determine which type of action is most effective, use some form of diagnostic indicators to measure the results of these actions (and assess the magnitude of the threats), and measure "early warning" indicators for other potential threats. If there are actions that would *clearly* abate the threats, the appropriate course of action would be to implement the actions at the scale needed to abate the threats, apply diagnostic indicators to measure the results of these actions, and measure "early warning" indicators for other potential threats.

Focusing on threats seems intuitively as though it would be a successful strategy for biodiversity conservation, with the potential to yield meaningful, measurable results in the short term. This focus may also provide managers with a valuable perspective on a different suite of indicators (threat variables) that could be selected for specific projects. Our review of the literature suggests that methods for measuring and quantifying threats to wildlife populations are perhaps less well developed than are methods for measuring status and response variables (which typically refer to species and habitats). The threat classification system developed by the Conservation Measures Partnership (IUCN-CMP 2006) represents a good step in this direction.

2. E.) Starting with "Known" Indicators: An example from the Adaptive Management literature

Adaptive management has been a part of the tool kit for natural resource managers for several decades and will no doubt be familiar to many readers. The core philosophy of adaptive management is "learning by doing," which means that management actions are regularly assessed to determine what worked and what didn't, and the lessons learned are applied to future projects or activities of a similar nature.

In a more formal context, adaptive management refers to a particular iterative management process which starts with the development of a logical model (often quantitative and computerized) of the system or process to be managed. The model predicts what will happen if particular management actions are taken, and includes specific indicators that are measured to determine whether or not the model's predictions are accurate. Information is collected during and after the implementation of management recommendations which is used to further refine the model, which will hopefully generate better predictions in the next management cycle.

A full discussion of adaptive management and associated modeling techniques is outside the scope of this review; interested readers are referred to more comprehensive treatments such as that of Walters (1986) and Stankey et al. (2005). However, for purposes of this review it is valuable to compare the role of indicators in the adaptive management cycle with the three approaches described above.

Walters (1986) describes a model-based approach for natural resource management. He recommends starting the process of model development by convening a working group that first identifies a few key management indicators and then builds a logical framework around the indicators. The framework includes all factors that might potentially influence each indicator. The process of building the framework continues until the project's management team feels that further elaboration is unnecessary. Other key indicators may be identified during the process of framework development. The completed framework is translated into a more formal, quantitative model (usually programmed in a computer), with the key indicators becoming the output variables from the model.

What is interesting here is that the selection of the key indicators takes place before the development of the conceptual model. Clearly, this is more easily done for certain natural resources such as fish stocks or timber reserves where there are a small suite of key measures such as population size or number of standing board feet that are generally accepted by the management community and would be measured as part of any resource management process. For many non-game wildlife populations and ecosystem or habitat types, it is much less clear what the most critical management metrics would be. In these cases, logic framework or causal chain approaches (which are both simple conceptual models) may be more helpful in identifying potential indicators.

The next section of this review will introduce the types of indicators and metrics that are commonly used in measuring the results of wildlife conservation activities.

3.) So… what do you actually measure?

This section provides a brief introduction to the theory of indicators and metrics, and then moves quickly into a discussion of the types of "things" that are commonly measured as part of monitoring and evaluation programs for wildlife management.

Theory: The Basic Types of Variables

Evaluation professionals distinguish among four basic types of variables or measures that form the basis for all indicators and metrics (Trochim 2006). Although this part of the discussion is somewhat technical, these variable types do matter, especially when one is attempting to aggregate or combine variables at larger spatial scales (statewide or nationally). The important thing to note here is that some types of variables add or combine easily, while others do not.

Nominal or **category** variables are names or descriptive labels attached to particular classes of objects. Different nominal variables cannot be added or subtracted. Vegetation type and species are both examples of nominal variables. It does not make sense to add iguanas and song sparrows, or mesic prairie and Sonoran desert scrub, for example! Note that the objects described by nominal variables can often also be described using other types of variables, such as ordinal (e.g. this site is the best example of Sonoran desert scrub) or ratio (e.g. 500 iguanas were seen in surveys).

Ordinal variables list quantities or items in some rank or order (smallest to largest, most rare to most common). While the ranks themselves are usually identified numerically (first, second, third, and so on) the distances between the numbers do not have any meaning (first place plus second place does not give you third place). Good examples of ordinal variables in the conservation field are the global, national, or state ranks (G-, N-, or S-ranks) in the NatureServe or Natural Heritage ranking system (NatureServe 2002).

Interval variables measure relative quantities such as humidity and temperature. Here the interval between measurements does have meaning – an interval of ten degrees Celsius always means the same thing at all points along the temperature scale, and is always greater than an interval of five degrees Celsius. It is possible to compute averages for interval variables (e.g. average temperature) but not ratios (since 50 degrees is not twice as hot as 25 degrees).

Ratio variables measure one quantity as a ratio of some other aspect of the environment (area, time, site, etc.). Most "count" variables are actually ratio variables where the denominator is one (one nature preserve, one state, one world). The number of tiger beetle species at a given nature preserve can be expressed as a ratio of the number of species to the number of preserves (in this case, one). The number of tiger beetle species in the world is a ratio of the number of species to the number of worlds (again, one). Composition, density, and concentration are other types of ratio measurements that are commonly used in wildlife management.

A distinction can also be made between **quantity** and **continuum** variables; quantity variables are those measured by ordinal numbers (1, 2, 3, etc.), while continuum variables can have values that are fractions of the ordinal numbers. Ordinal variables are quantity variables, while most interval and all ratio variables are continuum variables.

One final distinction often made in the evaluation literature is between **pressure**, **state**, and **response** variables (Organisation for Economic Co-operation and Development 1993). Pressure variables correspond to threats or stressors, state variables measure the current status of an environmental parameter, while response variables track the response of the environment to an action or other change. While this distinction may be helpful conceptually, authors such as Nielsen et al. (2001) have criticized it as too dependent on a particular theory of causality or environmental change that may or may not stand up under more rigorous investigation and analysis. According to these authors, if a manager bases the selection of indicators on a particular theory of causation and that theory changes, it may become necessary to change the indicators for a project (with the potential for the unpleasant discovery that one has been measuring the wrong thing all along). Regardless of whether one accepts these conclusions, it is important to point out that many of the same environmental variables would be measured as indicators of either state or response.

Things we actually measure

In summarizing our findings from the literature, we have found it useful to separate the "things" or attributes that are measured in wildlife management into two broad categories: **simple metrics** which are discrete units or things that are either directly measured in the field or use raw data from the field such as aerial photographs, and **composite metrics** which combine two or more simple metrics into a single rank or index value.

A more comprehensive list of potential indicators and metrics for specific types of management or restoration projects is provided in the Appendix. Readers are also referred to the "Further Reading" list for examples of published evaluation studies that use these and other metrics.

Simple metrics

Species population metrics: Given that many wildlife management projects are conducted for the purpose of recovering or improving populations of particular wildlife species, it is no surprise that there are a suite of widely-accepted metrics for evaluating the status of populations and species. Probably the simplest (at least conceptually) are the "count" variables (number of individuals, number of occurrences, number of populations or meta-populations). Counts per area can provide estimates of population density, while repeated counts over a series of time intervals can provide estimates of population trends.

For many species it is quite difficult to accurately census an entire population, so demographic or population models are used to estimate population size or population trends using data collected through statistically valid sampling schemes. The types of data that are actually collected from the wildlife population of interest will depend on the particular management model that is used, but such models often rely on data on reproductive output, age structure, survivorship, migration, and mortality rates. The Appendix lists a number of commonly collected demographic or population parameters.

Species composition: One of the simplest measures of biodiversity at a given site is the number of species (also known as species richness). There have been a number of more sophisticated metrics developed to measure biodiversity, but species richness remains popular due to the fact that it is easily calculated or estimated from field data. A manager may also be interested in the percentage of species at a site that share some particular ecological property, such as intolerance to disturbance. For instance, the percentage of ecologically sensitive macroinvertebrates is one of the individual metrics that contributes to the Index of Biotic Integrity for freshwater systems (discussed in more detail under "Composite Metrics" below).

Species distribution: Changes in the distribution or migratory patterns of species can provide powerful indirect evidence of significant changes in the local, regional, or global environment. At local scales, species distributions are typically quantified using element occurrence standards developed by NatureServe (2002). At larger spatial scales, species distributions are often displayed visually using range or point maps. Depending on the level of accuracy in these maps, changes in distribution may be measured quantitatively using GIS analysis.

Habitat extent: Most reporting of habitat protection and restoration activities is done using extent variables (number of acres protected, acquired, or restored; number of new miles of riparian forest buffer planted; size of conservation easement; number of miles of river opened to fish passage). Although easy to measure using modern surveying techniques, these variables have been roundly criticized in the literature (e.g. Ferraro and Pattanyak 2006) for providing little information about habitat quality or ecosystem processes that are critical for supporting wildlife populations.

Habitat composition: Composition metrics for wildlife habitat typically enumerate or describe aspects of the vegetative community (Ruiz-Jaen and Aide 2005). Such measures may be quantitative (numbers of floral species, numbers of canopy tree species), or qualitative (lists of the dominant species, as measured either by numbers of individuals or percent cover). At larger scales, habitat composition is usually measured by percent of particular land cover types within an area of interest (usually classified using either satellite or aerial imagery and some form of GIS analysis).

Habitat structure: These metrics describe physical parameters of the habitat itself – basal area of a forest stand, average height of vegetation in a grassland community, average height of understory shrubs, frequency of riffles and pools in a stream reach (Ruiz-Jaen and Aide 2005).

Habitat function/process: This is the hardest class of habitat variables to define, despite the fact that many ecologists argue it is one of the most important (e.g. Ruis-Jaen and Aide 2005). Part of the definitional complexity stems from the fact that many structural or compositional metrics also provide clues to ecosystem function. For example, seedling composition, height, and density together provide valuable information about forest stand recruitment and long-term stand dynamics. The depth of the soil organic layer is a structural attribute of soil, yet provides important insights into nutrient cycling. Another suite of variables commonly used in studying ecosystem function include concentrations of various chemicals (dissolved oxygen, nitrogen or phosphorous run-off, soil pH).

Some ecosystem processes have their own specialized measurement vocabulary and sets of associated indicators. For example, fire managers have developed their own sets of indicators and metrics to quantify pre- and post-fire fuel loads, burn extent, and burn frequencies (National Park Service 2003).

Resource variables: These variables describe key resources for wildlife species such as prey, water, host or food plants, and mutualistic partners (e.g. ants that tend the larvae of Karner Blue butterflies). Many resource variables are actually species or habitat measures, which would not otherwise be of management interest if they were not essential to particular wildlife species. For these variables, managers are typically interested in the presence, abundance, and spatial distribution of the resource.

Research and monitoring: Basic research is an important ancillary activity to wildlife conservation and is identified as a key activity in many of the state wildlife action plans. Research outputs are typically quantified in the academic sector by counting some combination of 1.) number of pages published in peer-reviewed journals, 2.) number of articles published in peer-reviewed journals, 3.) the impact factor of the journals in which the research is published or 4.) number of citations of a published article (Monastersky 2005).

Monitoring of wildlife populations is another important ancillary activity and there is an extensive literature on how to design monitoring programs to provide information at a level of detail sufficient to answer specific questions about species or populations of interest (Gibbs, Droege, and Eagle 1998). Some of the key questions in designing monitoring programs such as appropriateness of design, sensitivity, precision, and accuracy (Margoluis and Salafsky 1998), could also be used in evaluating these programs.

There has been particular interest in the use of volunteer or "citizen science" monitoring programs in recent years, and citations are provided below to published papers that review select citizen monitoring programs in more detail. In general, these studies are quite encouraging: with proper training and given simple tasks such as tree identification, citizen monitors achieved levels of accuracy comparable to those of more experienced field naturalists.

Regulatory Programs: Regulations that limit or prohibit hunting or collecting are one of the oldest and most widely used tools for managing wildlife, both game species as well as endangered species (Leopold 1933). While these programs make intuitive sense where hunting or collecting is a verifiable threat to the continued survival of wildlife populations, there have been few studies that have rigorously investigated the effects of these regulations. Available data are subject to interpretation: witness recent debates between advocates and critics of the federal Endangered Species Act which hinge on whether the number of species recovered (critics) or the number of species that have gone extinct (advocates) is an appropriate metric for judging the success of this act. In reviewing this debate, the Government Accountability Office (2006) argued that neither of these metrics provides a complete picture and that more information about the time and costs required for full achievement of recovery goals was needed to fairly evaluate the effectiveness of the Endangered Species Act.

Rodrigues (2006) suggests another metric which would be appropriate for regulatory programs as well as other species conservation activities: the number of species that would have gone extinct were it not for dedicated conservation actions. Obviously this metric would only be measurable for taxa such as birds where long-term population trend data are available.

Composite metrics

Composite metrics translate multiple quantitative measurements or qualitative assessments into a single metric that may facilitate comparisons between sites or across geographic levels. Such metrics are particularly helpful in combining measurements with different units (e.g. numbers of fish and average fish length) into a single metric for comparative purposes. Composite metrics may be composed of rank, interval, or ratio variables.

Probably the best known composite metric is the Index of Biotic Integrity for warmwater streams (Karr 1981), which in its original version combined 12 metrics (reflecting fish species richness and composition, number and abundance of indicator species, trophic organization and function, reproductive behavior, fish abundance, and condition of individual fish) into a single quantitative index scaled from 12 (lowest) to 60 (highest). The values of this index can be compared across stream segments, and the methodology can be also applied to different orders of streams, allowing some comparisons across different geographic scales.

The NatureServe (2002) or Natural Heritage ranking system is a series of nested ranked variables that measure the conservation status of rare species or unusual vegetation types at different geographic scales. It differs from the IBI in at least two important respects: the nested variables do not scale up according to a strict mathematical formula; and the system relies to a certain extent on expert judgment in establishing the rankings for each element at each level. Species or habitat occurrences are grouped using well-defined quantitative and/or qualitative standards into "Element Occurrences" or EOs. Within each state or territory, the rankings and number of EOs are used to establish a state or "S" rank. Within each country, the "S" ranks collectively help determine the national or "N" rank. And the various "N" ranks within a species' distribution help determine its global or "G" rank. (For species found in only one country, the "N" rank is usually synonymous with the "G" rank.)

A similar combination of metrics occurs in Red List indices, which measure changes in the global conservation status of groups of species which have been comprehensively assessed at least twice using the World Conservation Union (IUCN) methodology (Butchart et al. 2005). Red List indices are a very coarse assessment tool, but can nonetheless be used to demonstrate global changes in the conservation status of broad taxonomic categories such as birds or amphibians.

Non-biodiversity metrics

Wildlife conservation occurs in a social context and it is clear that factors in human society drive many of the processes that help or hinder wildlife populations. Understanding and describing these factors requires input from social scientists, economists, public health experts, and other evaluation professionals who may not always be included in discussions about wildlife conservation. An in-depth discussion of potential indicators and metrics for social processes is outside the scope of this review, but the interested reader is referred to standard textbooks such as Trochim (2006) for an introduction.

4.) Linking to the big picture – some metrics scale up, and some don't

This section is intended as an initial exploration of some of the practical issues associated with linking project- or program-specific performance metrics to national or global environmental indicators. A strong case can be made for attempting to develop such linkages. For practitioners on the ground, it would be highly beneficial to be able to show clear and compelling linkages between their specific local activities and the broader state of the national or global environment. For policy makers on the state or national level, as well as concerned citizens, it would also be valuable to have evidence that specific types of local actions are indeed contributing to environmental improvements.

Despite the potential practical importance of establishing these linkages, there have been few discussions of this topic in the literature, aside from a few early conceptual papers. Noss (1990) explored the concept of an integrated system for monitoring different levels of biodiversity, but suggested that different indicators may apply at different levels, a conclusion also reached by reviewers from The Wildlife Society (2002). While there is an extensive literature on related topics, such as the aggregation or disaggregation of spatial data, these discussions have been focused largely on statistical problems associated with the manipulation of particular types or classes of data. Although these topics are important and undoubtedly relevant to specific situations, the bigger question of whether it is even possible to create a national environmental reporting framework remains poorly explored.

It is unfortunate that so little attention has been paid to this issue, as there are considerable theoretical challenges associated with building an indicator framework that will enable local-level inputs to be linked to national indicators. Natural systems are extraordinarily complex and it is not clear in many circumstances which of many potential environmental variables should be measured to accurately assess system condition or project performance. There are also a vast number of potential "currencies" for measuring condition or performance, far more so than the analogous human economic system, for which indicators have been developed and refined over the past seventy years. And unlike the financial system, it is often difficult to see how one could make comparisons among the different types of environmental or ecosystem "currencies." Studies of the types of environmental or ecosystem indicators used by practitioners and decision-makers will undoubtedly help identify commonalities and potential common currencies on which an integrated indicator system can be built.

Can we even do this?

Much work has been done in recent years to develop both national environmental indicator systems (e.g. National Research Council 2000; The Heinz Center 2002; U.S. Environmental Protection Agency 2003) and local performance metrics (e.g. Margoluis and Salafsky 1998). Unfortunately there has been relatively little discussion of how these indicators and metrics might be linked into an integrated environmental reporting system. Rather, the questions driving these projects have included: "What are the suite of indicators that best describe environmental condition in the United States?" and "What can I measure to tell if my project is having the desired environmental results?" (Heinz Center 2002; Stem et al. 2005). When the questions are posed in this way, it can become difficult to see linkages between local performance measurement and national environmental assessment.

Recent authors (e,g, Salzer and Salafsky 2006; Stem et al. 2005) have drawn a clear distinction between two different types of monitoring and evaluation activities: status assessments, which attempt to determine the condition, state, or trend of one or more variables in the environment; and effectiveness measurements, which track changes associated with the implementation of particular management activities. If we follow this approach, then existing national environmental indicators (such as those contained in the "State of the Nation's Ecosystems") would be seen as a type of status assessment, while project- or program- specific metrics would mostly be classified as effectiveness measurements. Stem et al. (2005) use these two categories to classify various evaluative approaches and tools, leading to the conclusion that different approaches are needed for each type of assessment activity.

While there may be important methodological distinctions between status assessment and effectiveness measurement, it is important to note that these two activities often look at similar – or even identical – environmental variables. For example, a manager for a tallgrass prairie site in Iowa may track populations of grassland-dependent sparrows to evaluate the effectiveness of specific management practices such as prescribed fire. At the same time, population numbers of these birds directly inform two of the national indicators in the "State of the Nation's Ecosystems" report (Heinz Center 2002), and also inform the global "Red List" indicator for bird species (Butchart et al. 2005).

Similarities between the types of things that are being monitored or measured suggest that there is value to exploring links between status assessments and effectiveness measures.

Some key concepts

Levels: There are at least two informal hierarchies of scale that can be used to describe conservation actions and environmental conditions. Each hierarchy has several levels. One hierarchy is a scale of effort (for which some common levels include activity, project, and program); the other is geographic scale (for which common levels include local, regional, state, national, and global). Performance measurements and indicators are usually designed to answer questions at a particular level in one or both of these hierarchies. It is important to note that except for some of the higher levels (e.g. state, national, global), boundaries between levels are "fuzzy" - one person's program may be another's project; one group's regional boundaries may be local to another group. And of course the boundaries between different "regions" at the same scale may be similar but non-identical (witness the different regional structures developed by the federal land management agencies).

Aggregation: Combining measurements or metrics on one level into a single metric (which may or may not be at a higher level). For example, adding the total farm acreages for each state gives you the total farm acreage for the country.

Disaggregation: Breaking apart a large data set into one or more data sets with smaller resolution (which may or may not be at a lower level).

When does it work?

The remainder of this section will apply a "case study" approach to investigate different types of indicators that are commonly used in conservation practice and whether – or not – aggregation or disaggregation between levels is even possible.

Examples of successful aggregation

1.) Same units used at both local and national levels; national metric is the sum of local measurements

An obvious example of this is the "no net loss" national goal for wetlands in the United States. Here, an acre of wetlands conserved at the local scale contributes directly to the national accounting of wetland acreage, which in turn determines whether or not the overall national goal has been met.

Another fairly obvious example is in the area of endangered species biology and management. For the federally endangered whooping crane (as with many endangered species), recovery goals are defined in terms of the number of birds in the various migratory, non-migratory, and captive populations. The same units (number of birds) are used at the local level, the population level, and at the national level in assessing the recovery status of the species as a whole.

2.) Indices and nested indices.

As discussed above, indices translate multiple quantitative measurements into a single metric that may facilitate comparisons between sites or across geographic levels. Probably the best known index is the Index of Biotic Integrity for warmwater streams (Karr 1981), described in more detail above. The values of this index can be compared across stream segments, and the methodology can be also applied to different orders of streams, allowing some comparisons across different geographic scales.

The NatureServe or Natural Heritage ranking system is a series of nested indices that measure the conservation status of rare species or unusual vegetation types at different geographic scales. The ranking given to a species or vegetation type at a lower scale (e.g. state) helps to inform the ranking for that same species or vegetation type at a higher scale (e.g. national or global). The NatureServe ranking system differs from the IBI in at least two important respects: the nested indices do not scale up according to a strict mathematical formula; and the system relies in part on expert judgment in establishing the rankings for each element at each level.

The "S", "N", or "G" rankings in the NatureServe system can also be pooled for particular taxa or for a particular geographic area to give "snapshots" of the state of biodiversity at a particular level. For the Heinz Center's "State of the Nation's Ecosystems" report (Heinz Center 2002), the national indicator of at-risk native species reports the percentage of U. S. species of plants and animals that have been ranked by NatureServe as either "vulnerable," "imperiled," or "critically imperiled" globally.

An interesting property of indices is that, while they may facilitate comparisons at the same or higher levels, it may not be possible to disaggregate the higher-level indices into the component metrics. To give a simple example using the NatureServe ranking system, two species may both have a global rank of "vulnerable" (G3 in technical terms), but the lower-level "S ranks" within the individual states where this species occurs cannot be predicted from the global rank information alone. One of the species may have a few large occurrences in a few states, while the other species may have many smaller occurrences across numerous states.

What are the limits to aggregation / disaggregation?

1.) Nominal variables and ordinal variables

We discussed the four basic types of variables above. Nominal variables such as "ecosystem type" cannot be added or aggregated – one cannot combine, say, a desert and a tundra into something meaningful. One can, of course, add ratio variables such as area that describe nominal variables such as ecosystem type. Ordinal variables cannot be added or averaged (a G1 plus a G2 does not equal a G3), while interval variables such as temperature measurements can be averaged but not added. In contrast, most standard mathematical operations can be performed on ratio variables. Of course, there are other rules that will govern whether or not the results of any particular mathematical operation performed on ratio variables are meaningful from a biological or ecosystem perspective.

2.) Area-dependent metrics

Some metrics are tightly tied to particular spatial scales. Density measurements – the number of trees, animals, etc. per unit area – are a good example. When working with natural systems, it is usually not possible to compare density measurements made at one scale with density measurements of the same objects at a different scale. This is because natural objects usually have non-uniform distributions, and changing the unit area of measurement may yield larger or smaller densities depending on the underlying distribution of the individuals, objects, etc. of interest.

3.) Minimum resolution and spatial data sets

Satellite or GIS data sets frequently have a specified minimum resolution. The use of one of these data sets usually sets a lower bound on disaggregation. For example, The Heinz Center's new landscape pattern indicator uses a minimum resolution of 30 square meters. While it is possible to combine these "pixels" of data into a data set with larger pixels, it is not possible to disaggregate these data into pixels smaller than 30 square meters.

There are special cases where disaggregation is possible if the behavior of the variables in the data set is well known. For example, a low-resolution temperature map can be disaggregated through interpolation with a high-resolution digital elevation model, because the relationship between elevation and temperature is well understood.

4.) Local responses not detectable at global or national scales

A significant difficulty for natural resource managers is the fact that the effects of individual management projects may not be detectable using metrics designed for national or global scales. Many projects are implemented at very small geographic scales and the positive benefits from these projects are swamped by other environmental factors. Even when many smaller projects are aggregated, their collective effects may not be detectable. For example, conservationists have planted 2000+ miles of new riparian forest buffer within the Chesapeake Bay watershed since 2000, yet there has been a notable lack of improvement in nutrient levels in the mainstem of Chesapeake Bay, due to overwhelming nutrient inputs from urban, suburban, and agricultural sources.

Given the limits of resolution for many widely available satellite or GIS data sets, even the aggregate effects of numerous projects may not be perceived at such large scales. To continue with the Chesapeake Bay example, the 2000+ miles of riparian forest plantings may not even be detectable in the USDA Forest Service's Forest Inventory and Analysis (FIA) National Program, which is the most comprehensive assessment of forest ecosystem changes in the U. S. A. This is due to the fact that FIA utilizes Landsat Thematic Mapper imagery, which has a minimum resolution of 30 meters, and many of the individual riparian buffer plantings are smaller than 30 meters in one or more dimensions.

5.) Difficulties with aggregation

Some data sets are very difficult to aggregate (e.g. USGS groundwater sampling wells, where the data collection points are far apart and there are few intuitive ways to combine or aggregate these data). In these cases it may be necessary to build a quantitative spatial model to generate the spatial contiguity necessary for aggregation/disaggregation operations. In especially difficult cases where there are significant gaps in our sampling or in our understanding of key ecosystem processes, it may not even be possible to construct a model that would provide a meaningful context for the data. And of course models have their own set of problems, mainly due to the assumptions and simplifications which are embedded in them.

Some indicators don't aggregate or disaggregate – is that okay?

Not all local performance metrics will aggregate into one of the national or global environmental indicators. This is especially true if one is using a national indicator set such as that found in the Heinz Center's State of the Nation's Ecosystem report, where a conscious effort has been made to identify a relatively small group of indicators. By no means does this imply that other metrics are not valuable or non-essential. For example, it is critically important for managers to measure survivorship among tree seedlings in a riparian buffer planting, yet it is difficult to imagine how that measurement might be aggregated into a regional or national metric. Number of eggs per clutch is an important measure of the health of certain bird populations, yet again it is difficult to see how this metric would translate to a national scale.

It is also important to note that many larger scale metrics do not scale down. Just because average global temperature changes by a few degrees does not mean that the temperature at any given location will change by the same number of degrees. The percentage of the U. S. land base that is currently forested does not provide information on the amount of land that is forested in the state of Idaho. And the fact that a certain percentage of the nation's species are at risk for extinction does not give information on the status of any one of these species.

5.) Further Reading

1.) Why Measure?

Fazey, I., J. G. Salisbury, D. B. Lindenmayer, J. Maindonald, and R. Douglas. 2004. Can methods applied in medicine be used to summarize and disseminate conservation research? Environmental Conservation 31(3):190-198.

Healy, W. M. (ed.) 2002. Performance Measures for Ecosystem Management and Ecological Sustainability. The Wildlife Society Technical Review 02-1:i-vi + 1-33.

Kleiman, D. G., R. P. Reading, B. J. Miller, T. W. Clark, J. M. Scott, J. Robinson, R. L. Wallace, R. J. Cabin and F. Felleman. 2000. Improving the evaluation of conservation programs. Conservation Biology 14(2):356-365.

Margoluis, R., and N. Salafsky. 1998. Measures of Success: Designing, Managing, and Monitoring Conservaiton and Development Projects. Island Press, Washington, D. C. xx + 362 pp.

Noon, B.R. 2003. Conceptual issues in monitoring ecological resources. Pages 27-72 In: Monitoring Ecosystems: Interdisciplinary Approaches for Evaluating Ecoregional Initiatives. D.E. Busch and J.C. Trexler (eds.). Island Press.

Pullin, A. S. and T. M. Knight. 2001. Effectiveness in conservation practice: Pointers from medicine and public health. Conservation Biology 15(1):50-54.

Salafsky, N., R. Margoluis, K. H. Redford and J. G. Robinson. 2002. Improving the practice of conservation: A conceptual framework and research agenda for conservation science. Conservation Biology 16(6):1469-1479.

Stankey, G. H., R. N. Clark, and B. T. Bormann. 2005. Adaptive Management of Natural Resources: Theory, Concepts, and Management Institutions. USDA Forest Service Pacific Northwest Research Station General Technical Report PNW-GTR-654. 73 pp.

Stem, C., R. Margoluis, N. Salafsky and M. Brown. 2005. Monitoring and evaluation in conservation: A review of trends and approaches. Conservation Biology 19(2):295-309.

The Wildlife Society. 2002. Performance Measures for Ecosystem Management and Ecological Sustainability. The Wildlife Society Technical Review 02-1. vi + 33 pp.

Trochim, W. M. K. 2006. Research Methods Knowledge Base. World Wide Web site at: http://www.socialresearchmethods.net/kb/index.php

Walters, C. 1986. Adaptive Management of Renewable Resources. The Blackburn Press, Caldwell, New Jersey. x + 374 pp.

2.) How Do You Develop Measures?

Hagan, J. M. and A. A. Whitman. 2004. A Primer on Selecting Biodiversity Indicators for Forest Sustainability: Simplifying Complexity. Forest Mosaic Science Notes FMSN-2004-1:1-4.

Salafsky, N., R. Margoluis, K. H. Redford and J. G. Robinson. 2002. Improving the practice of conservation: A conceptual framework and research agenda for conservation science. Conservation Biology 16(6):1469-1479.

Salzer, D. and N. Salafsky. 2006. Allocating resources between taking action, assessing status, and measuring effectiveness of conservation actions. Natural Areas Journal 26(3):310-316.

Tear, T. H., P. Kareiva, P. L. Angermeier, P. Comer, B. Czech, R. Kautz, L. Landon, D. Mehlman, K. Murphy, M. Ruckelshaus, J. M. Scott and G. Wilhere. 2005. How much is enough? The recurring problem of setting measurable objectives in conservation. BioScience 55(10):835-849.

Walters, C. 1986. Adaptive Management of Renewable Resources. The Blackburn Press, Caldwell, New Jersey. x + 374 pp.

3.) So… What Do You Actually Measure?

Indicator Theory

Organisation for Economic Co-operation and Development. 1993. OECD core set of indicators for environment performance reviews. World Wide Web document at: http://www.oecd.org/env/docs/gd93179.pdf

Trochim, W. M. K. 2006. Research Methods Knowledge Base. World Wide Web site at: http://www.socialresearchmethods.net/kb/index.php

Species

Cantu, R. and C. Richardson. 1997. Mule deer management in Texas. Texas Parks and Wildlife Department. 26 pp.

Evans, J. E., W. N. Grafton, and T. R McConnell. 1999. Fundamentals of deer harvest management. West Virginia Division of Natural Resources Cooperative Extension Service Publication 806:1-4.

Gibbs, J. P., S. Droege, and P. Eagle. 1998. Monitoring populations of plants and animals. BioScience 48(11):935-940.

Nielsen, J. R., P. Degnbol, H. Hovgaard, and S. Reeves. 2001. Indicators as a basis for robust and acceptable fisheries management. In: Proceedings of the Regional Technical Consultation on Indicators for Sustainable Fisheries Management in ASEAN Region. 2-5 May 2001, Haiphong, Vietnam. 60-70.

Niemi, G. J., J. M. Hanowski, A. R. Lima, T. Nicholls, and N. Weiland. 1997. A critical analysis on the use of indicator species in management. Journal of Wildlife Management 61(4):1240-1252.

Ralls, K. and J. D. Ballou. 2004. Genetic status and management of California Condors. The Condor 106:215-228.

Sharma, R., G. Morishima, S. Wang, A. Talbot, and L. Gilbertson. 2006. An evaluation of the Clearwater River supplementation program in western Washington. Canadian Journal of Fisheries and Aquatic Science 63:423-437.

Shult, M. J., and B. Armstrong. 1999. Deer census techniques. Texas Parks and Wildlife 4 pp.

Habitat Protection

Bay Area Open Space Council. 1999. Ensuring the promise of conservation easements: Report on the use and management of conservation easements by San Francisco Bay area organizations. Bay Area Open Space Council, San Francisco, CA. 46 pp.

Bruner, A. G., R. E. Gullison, R. E. Rice, and G. A. B. de Fonseca. 2001. Effectiveness of parks in protecting tropical biodiversity. Science 291:125-128.

Ferraro, P. J. and S. K. Pattanayak. 2006. Money for nothing? A call for empirical evaluation of biodiversity conservation investments. Public Library of Science: Biology 4(4):e105. (online at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1435411).

Land Trust Alliance. 1999. How strong are our defenses: The results of the Land Trust Alliance's northern New England conservation easement quality research project. Land Trust Alliance, Washington, D.C. 63 pp.

Habitat Restoration

General

Anand, M. and R. E. Desrochers. 2004. Quantification of restoration success using complex systems concepts and models. Restoration Ecology 12(1):117-123.

Bell, S. S., M. S. Fonseca and L. B. Motten. 1997. Linking restoration and landscape ecology. Restoration Ecology 5(4):318-323.

Block, W. M., A. B. Franklin, J. P. Ward, Jr., J. L. Ganey and G. C. White. 2001. Design and implementation of monitoring studies to evaluate the success of ecological restoration on wildlife. Restoration Ecology 9(3):293-303.

Ehrenfeld, J. G. 2000. Defining the limits of restoration: The need for realistic goals. Restoration Ecology 8(1):2-9.

Hobbs, R. J. and D. A. Norton. 1996. Towards a conceptual framework for restoration ecology. Restoration Ecology 4(2):93-110.

Michener, W. K. 1997. Quantitatively evaluating restoration experiments: Research design, statistical analysis, and data management considerations. Restoration Ecology 5(4):324-337.

Ruiz-Jaen, M. C., and T. M. Aide. 2005. Restoration success: How is it being measured? Restoration Ecology 13(3):569-577.

Smallwood, K. S. 2001. Linking habitat restoration to meaningful units of animal demography. Restoration Ecology 9(3):253-261.

White, P. S. and J. L. Walker. 1997. Approximating nature's variation: Selecting and using reference information in restoration ecology. Restoration Ecology 5(4):338-349.

Streams and Rivers

Brooks, Shane S., M. A. Palmer, B. J. Cardinale, C. M. Swan, and S. Ribblett. 2002. Assessing stream ecosystem rehabilitation: Limits of community data structure. Restoration Ecology 10(1):156-168.

Frissell, C. A. and R. K. Nawa. 1992. Incidence and causes of physical failure of artificial habitat structures in streams of western Oregon and Washington. North American Journal of Fisheries Management 12:182-197.

Karr, J. R. 1981. Assessment of biotic integrity using fish communities. Fisheries (Bethesda) 6: 21-27.

Kondolf, G. M. 1995. Five elements for effective evaluation of stream restoration. Restoration Ecology 3(2):133-136.

Kondolf, G. M., J. C. Vick, and T. M. Ramirez. 1996. Salmon spawning habitat rehabilitation on the Merced River, California: An evaluation of project planning and performance. Transactions of the American Fisheries Society 125:899-912.

Palmer, M. A., E. S. Bernhardt, J. D. Alla, P.S. Lake, G. Alexander, S. Brooks, J. Carr, S. Clayton, C. N. Dahm, J. Follstad Shah, D. L. Galat, S. G. Loss, P. Goodwin, D.D. Hart, B. Hassett, R. Jenkinson, G.M. Kondolf, R. Lave, J.L. Meyer, T.K. O'Donnell, L. Pagano and E. Sudduth. 2005. Standards for ecologically successful river restoration. Journal of Applied Ecology 42:208-217.

Purcell, A. H., C. Friedrich and V. H. Resh. 2002. An assessment of a small urban stream restoration project in northern California. Restoration Ecology 10(4):685-694.

Wetlands

Faber-Langendoen, D., J. Rocchio, M. Schafale, C. Nordman, M. Pyne, J. Teague, T. Foti, and P. Comer. 2006. Ecological Integrity Assessment and Performance Measures for Wetland Mitigation. NatureServe, Arlington, Virginia. 44 pp.

Lichko, L. E. and A. J. K. Calhoun. 2003. An evaluation of vernal pool creation projects in New England: Project documentation from 1991–2000. Environmental Management 32(1):141-151.

Neckles, H. A., M. Dionne, D. M. Burdick, C. T. Roman, R. Buchsbaum, E. Hutchins. 2002. A monitoring protocol to assess tidal restoration of salt marshes on local and regional scales. Restoration Ecology 10(3):556-563.

Windham, L., M. S. Laska, and J. Wollenberg. 2004. Evaluating urban wetland restorations: Case studies for assessing connectivity and function. Urban Habitats 2(1):130-146.

Grasslands

Packard, S. and C. F. Mutel. The tallgrass restoration handbook: For prairies, savannas, and woodlands. Island Press: Washington, D. C. 495 pp.

Shepard, S., and D. M. Debinski. 2005. Evaluation of isolated and integrated prairie reconstructions as habitat for prairie butterflies. Biological Conservation 126:51-61.

Forests

King, S. L. and B. D. Keeland. 1999. Evaluation of reforestation in the lower Mississippi alluvial valley. Restoration Ecology 7(4):348-259.

Parkyn, S. M., R. J. Davies-Colley, N. J. Halliday, K. J. Costley, and G. F. Croker. 2003. Planted riparian buffer zones in New Zealand: Do they live up to expectations? Restoration Ecology 11(4):436-447.

Stanturf, J. A., S. H. Schoenholtz, C. J. Schweitzer, and J. P. Shepard. 2001. Achieving restoration success: Myths in bottomland hardwood forests. Restoration Ecology 9(2):189-200.

Function/Process

Ruiz-Jaen, M. C., and T. M. Aide. 2005. Restoration success: How is it being measured? Restoration Ecology 13(3):569-577.

National Park Service. 2003. Fire Monitoring Handbook. Boise (ID): Fire Management Program Center, National Interagency Fire Center. 274p.

Research

Monastersky, R. 2005. The number that's devouring science. The Chronicle of Higher Education, October 14, 2005. Available online at: http://chronicle.com/free/v52/i08/08a01201.htm

Monitoring

Danielsen, F., N. D. Burgess, and A. Balmford. 2005. Monitoring matters: examining the potential of locally-based approaches. Biodiversity and Conservation 14:2507-2542.

Foster-Smith, J. and S. M Evans. 2003. The value of marine ecological data collected by volunteers. Biological Conservation 113:199-213.

Newman, C., Buesching, C. D., and Macdonald, D. W. 2003. Validating mammal monitoring methods and assessing the performance of volunteers in wildlife conservation – "*Sed quis custodiet ipsos custodies*?" Biological Conservation 113:189-197.

Regulatory Programs

Leopold, A. 1933. Game Management. New York: Charles Scribner's Sons. 479 pp.

Rodrigues, A. S. L. 2006. Are global conservation efforts successful? Science 313:1051-1052.

U. S. Government Accountability Office. 2006. Endangered Species: Time and Costs Required to Recover Species Are Largely Unknown. GAO-06-463R:1-29.

Composite metrics

Butchart, S. H. M., A. J. Stattersfield, J. Baillie, L. A. Bennun, S. N. Stuart, H. R. Açkakaya, C. Hilton-Taylor, and G. M. Mace. 2005. Using Red List Indices to measure progress towards the 2010 target and beyond. Philosophical Transactions of the Royal Society B 360:255-268.

Karr, J. R. 1981. Assessment of biotic integrity using fish communities. Fisheries (Bethesda) 6: 21-27.

NatureServe. 2002. Element Occurrence Data Standard. NatureServe, Arlington, VA. 201 pp.

Salzer, D. and N. Salafsky. 2006. Allocating resources between taking action, assessing status, and measuring effectiveness of conservation actions. Natural Areas Journal 26(3):310-316.

4.) Linking to the big picture

Butchart, S. H. M., A. J. Stattersfield, J. Baillie, L. A. Bennun, S. N. Stuart, H. R. Açkakaya, C. Hilton-Taylor, and G. M. Mace. 2005. Using Red List Indices to measure progress towards the 2010 target and beyond. Philosophical Transactions of the Royal Society B 360:255-268.

Karr, J. R. 1981. Assessment of biotic integrity using fish communities. Fisheries (Bethesda) 6: 21-27.

National Research Council, Committee to Evaluate Indicators for Monitoring Aquatic and Terrestrial Environments. 2000. Ecological indicators for the nation. National Academies Press, Washington, D. C. 198 pp.

NatureServe. 2002. Element Occurrence Data Standard. NatureServe, Arlington, VA. 201 pp.

Noss, R. F. 1990. Indicators for Monitoring Biodiversity: A Hierarchical Approach. Conservation Biology 4(4):355-264.

Salzer, D. and N. Salafsky. 2006. Allocating resources between taking action, assessing status, and measuring effectiveness of conservation actions. Natural Areas Journal 26(3):310-316.

Stem, C., R. Margoluis, N. Salafsky and M. Brown. 2005. Monitoring and evaluation in conservation: A review of trends and approaches. Conservation Biology 19(2):295-309.

The Heinz Center. 2002. State of the Nation's Ecosystems. Cambridge University Press, Cambridge, U. K. 270 pp.

The Wildlife Society. 2002. Performance Measures for Ecosystem Management and Ecological Sustainability. The Wildlife Society Technical Review 02-1. vi + 33 pp.

U. S. Environmental Protection Agency. 2003. Draft report on the environment. World Wide Web document at: http://www.epa.gov/indicators/roe/html/roeTOC.htm

Acknowledgments

For discussions on these subjects, we thank Dennis Murphy of University Nevada-Reno; Matthew Birnbaum, Brian Gratwicke, Tom Kelsch and Peter Stangel of the National Fish and Wildlife Foundation; Richard Margoluis and Nick Salafsky of Foundations of Success; Andy Rowe of GHK, Inc.; David Chadwick and Naomi Edelson of the Association of Fish and Wildlife Agencies; Michael Hickey of the White House Office of Management and Budget; Karen Terwilliger of Terwilliger Consulting, Inc.; Genevieve Larouche and Greg Watson from the U. S. Fish and Wildlife Service; Gaby Chavarria of the Natural Resources Defense Council; Jeff Lerner and Sara Vickerman at Defenders of Wildlife, Corinne Carter of NASA Goddard Space Flight Center; the state representatives who attended the Wildlife Diversity Managers Meeting in Warner Springs, California, in January, 2007; members of the Teaming with Wildlife coalition who attended the 72nd North American Wildlife and Natural Resources Conference in Portland, Oregon; and our Heinz Center colleagues Tom Lovejoy, Christophe Tulou, Kent Cavender-Bares, Christine Negra, Anne Marsh, Caroline Cremer, Holly MacCormick, and Claire Hayes.

Appendix 1
Some common variables used in monitoring and evaluating wildlife management programs

SPECIES MANAGEMENT

- "Count" variables
  - Number of animals/plants
  - Number of animals/plants per unit area
  - Time series of population numbers
  - Abundance (including trends)
- Composition
  - Number of Species (global)
  - Number of Species per unit area
  - Percentage of a sample/area that is some particular taxon
- Geographic distribution
  - Range maps (and changes to them)
  - Density estimates (distribution of entities across landscapes)
  - Movement/migratory patterns
- Basic life history or demographic parameters
  - Dimensional measurements (length, antler size) incl. maximum size
  - Growth rate of individual organisms
  - Life history chronology
  - Age or Age class distribution
  - Weight or Weight class distribution
  - Survivorship/Mortality
  - Reproductive output
  - Population growth rate
  - Sex ratios
  - Genetic stability in populations
  - Genetic diversity in populations
  - Disease prevalence
- Resources
  - Food
    - Availability
    - Density
    - Production
  - Water
    - Availability
  - Light
    - Availability
    - Heterogeneity
  - Other key species (e.g. ants that tend Karner Blue butterfly caterpillars)
    - Availability

HABITAT EXTENT VARIABLES

- Absolute size of protected area
- Relative size of protected area
- Number of protected areas
- Biome and habitat content of protected area
- Species composition of protected area
- Degree of connectivity/separation from other protected areas
  - Landscape composition
    - Percent land use
    - Percent land in “natural condition”
    - Percent land unfragmented within a given distance (e.g. 1 km)
  - Immediately adjacent land use to protected area
  - Distance to nearest road / road network extent
  - Connectivity to other habitat patches
  - Continuity of riparian corridor
  - Area of contiguous fire-maintained landscape

HABITAT MANAGEMENT (most structural, some compositional or functional)

General

- Biomass
- Floristic quality assessment
- Vegetation Index of Biotic Integrity
- Invasive species (presence/absence, number, percent cover of plants)
- Presence of particular indicator species for ecosystem/habitat type

Soils

- pH of soil water
- organic soil horizons
- soil organic matter decomposition
- soil organic carbon
- soil bulk density

Forest

- Basal area
- Percent canopy closure
- Other metrics of vegetative structure (height of understory, canopy, supercanopy)
- Density of pole trees
- Density of regenerating trees
- Presence/amount of coarse woody debris

Grassland

- Amount of litter
- Depth of litter
- Invasives (species list, percent cover, other density metrics)
- Woody species recruitment
- Woody fuel loading (per area)
- Mean height-density obstruction
- Mean height-disc readings per field

Streams and Rivers

Index of Biotic Integrity (original version [note that most metrics that comprise IBI are “count” or composition metrics])
Channel depth
Substrate size
Substrate embeddedness
Velocity/depth
Water clarity
Water temperature
Water chemistry metrics (see below)
Sediment deposition
Channel flow status
Channel alteration
Frequency of riffles
Bank stability
Condition of buffer
Presence/absence of coarse woody debris

Wetlands

Water clarity
Water temperature
Water chemistry metrics (see below)
Upstream surface water retention
Upstream/onsite water diversions
Flashiness index
Floodplain interaction
Water table depth
Surface water runoff index
Hydrological alterations
Presence/absence of coarse woody debris
Biotic patch richness
Interspersion of biotic patches
Presence/absence of beaver activity
Litter cover

FUNCTION OR PROCESS VARIABLES

(Note: Many of the structural variables listed above under “Habitat Management” may also provide valuable insights into ecosystem function.)

Concentration of some nutrient/chemical

Nitrogen concentration
Phosphorous concentration
Dissolved oxygen concentration
pH of soil, water

Fire

Mean abundance (biomasss or percent cover) of the dominant species at the site
Historic ecosystem fire regime
- Fire severity
- Fire return interval

Recovery time following fire