A Framework for Improved Monitoring of Biodiversity: Responses to the World Summit on Sustainable Development

RHYS E. GREEN,*†‡‡ ANDREW BALMFORD,* PETER R. CRANE,‡ GEORGINA M. MACE,§ JOHN D. REYNOLDS,** AND R. KERRY TURNER††

*Conservation Biology Group, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, United Kingdom

†Royal Society for the Protection of Birds, The Lodge, Sandy, Bedfordshire SG19 2DL, United Kingdom

‡Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB, United Kingdom

§Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY, United Kingdom

**Centre for Ecology, Evolution and Conservation, School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, United Kingdom

††Centre for Social and Economic Research on the Global Environment, University of East Anglia, Norwich NR4 7TJ, United Kingdom

Abstract: *The Convention on Biological Diversity (CBD) and the World Summit on Sustainable Development (WSSD) endorsed the Hague Ministerial Declaration that calls for a significant reduction of the current rate of biodiversity loss at the global, regional, and national levels by 2010. We argue that there is a shortage of standardized, regularly repeated measurements of the state of biomes and their biota that could be used to monitor progress toward this goal. In particular, there are few data that directly or indirectly measure the delivery of ecosystem services that depend on biodiversity. Given the link made in the declaration between biodiversity and poverty alleviation, this deficiency is of special concern. We suggest that greater attention should be given to defining the questions about changes in biodiversity that are relevant to CBD and WSSD goals and propose a framework through which the links between these questions and programs of monitoring and research could be made stronger and more explicit. The framework consists of three stages. First is a scoping stage in which reviews of existing knowledge and interactions with stakeholders help to define the subject of the evaluation and lead to a preliminary model of the system of interest. Second is a design stage in which the types of measurement and sampling strategies are selected by evaluating their fitness for purpose and the resources available to conduct the work. The final stage is implementation and reporting, which considers data collection and storage and the evaluation and dissemination of results. This framework can be applied across a broad range of biodiversity attributes and scales and, if combined with a systematic review of the most important and relevant questions about changes in biodiversity, would improve the coverage, fitness for purpose, and value for money of biodiversity monitoring. Slowing the rate of loss of biodiversity requires conservation action, but to know where this is most needed and whether it is working requires better and more comprehensive monitoring.*

Key Words: assessment, biodiversity measurement, biodiversity monitoring, evaluation, World Summit on Sustainable Development

Un Marco para Mejorar el Monitoreo de Biodiversidad: Respuestas a la Cumbre Mundial de Desarrollo Sustentable

Resumen: *La Convención de Diversidad Biológica (CDB) y la Cumbre Mundial de Desarrollo Sustentable (CMDS) avalaron la Declaración Ministerial de La Haya que llama a una reducción significativa de la tasa*

‡‡*email reg29@hermes.cam.ac.uk*

Paper submited June 25, 2003; revised manuscript accepted June 24, 2004.

actual de pérdida de biodiversidad en los niveles global, regional y nacional para 2010. Argumentamos que hay escasez de mediciones estandarizadas, repetidas regularmente, del estado de biomas y su biota que pudieran ser utilizadas para monitorear el progreso hacia esa meta. En particular, hay pocos datos que miden, directa o indirectamente, el reparto de servicios ecosistémicos que dependen de la biodiversidad. Dada la relación entre la declaración entre biodiversidad y disminución de pobreza, esta deficiencia es de especial preocupación. Sugerimos se debe dar mayor atención a definir preguntas sobre cambios en la biodiversidad que son relevantes para metas de CDB y CMDS y proponer un marco en el que se pueden reforzar y hacer más explícitos los enlaces entre estas preguntas y programas de monitoreo e investigación. El marco consiste de tres etapas. La primera es una etapa de visualización en la que se revisa el conocimiento existente y se interactúa con los actores para definir el sujeto de evaluación y desarrollar un modelo preliminar del sistema de interés. La segunda es la etapa de diseño en la que se seleccionan estrategias de medición y muestreo así como los recursos disponibles para desarrollar el trabajo. La etapa final es la implementación y reporte, que considera la recolecta de datos y el almacenamiento y la evaluación y diseminación de resultados. Este marco puede aplicarse en una amplia gama de atributos y escalas de biodiversidad y, si se combina con una revisión sistemática de las preguntas más importantes y relevantes sobre cambios de biodiversidad, podría mejorar la cobertura, adecuación y valor del monitoreo de biodiversidad. Disminuir la tasa de pérdida de biodiversidad requiere de acciones de conservación, pero para saber donde se necesita y si esta funcionando requiere de monitoreo más adecuado e integral.

Palabras Clave: Cumbre Mundial de Desarrollo Sustentable, estimación, evaluación, medición de biodiversidad, monitoreo de biodiversidad

Introduction

The Convention on Biological Diversity (CBD) recognizes the importance to sustainable development of protecting biological diversity and functioning ecosystems (Secretariat of the Convention on Biological Diversity 2003). Biodiversity provides a wide range of significant use and nonuse benefits and essential life-support services (Costanza et al. 1997; Daily 1997; Balmford et al. 2002). In its Plan of Implementation, the 2002 World Summit on Sustainable Development (WSSD) endorsed the Hague Ministerial Declaration of the Sixth Conference of the Parties to the CBD that committed them "to achieve by 2010 a significant reduction of the current rate of biodiversity loss at the global, regional and national level as a contribution to poverty alleviation and to the benefit of all life on earth." (United Nations 2002). If governments and intergovernmental organizations are to achieve progress toward this objective, they will need to measure how well they are doing. This will require a series of standardized, regularly repeated measurements of the state of biomes and their biota. Given the link made in the declaration between biodiversity and poverty alleviation, these measurements must capture information on the area of biomes; the diversity, distribution, and abundance of species; and the provision of ecosystem goods and services. It is clear, however, that there is a great deal of work to do before the scientific community can generate reliable time series of data relevant to this target for a wide range of biomes and attributes at the global level.

There are a number of initiatives to assess the state of the world's biodiversity. The most prominent program is the Millennium Ecosystem Assessment (2004), which is currently reviewing the status and trends of the world's biodiversity primarily from the standpoint of ecosystem services (Reid & Mace 2003). This is a timely and hugely ambitious initiative, but it is based nearly entirely on existing information about global biodiversity at one point in time; thus, it is principally an assessment rather than an ongoing monitoring program.

In terms of repeated measurements, there are a handful of regional or national monitoring programs that measure a broad range of components of biodiversity and ecosystem services—most notably the Heinz Center's program for the United States (The H. John Heinz III Center for Science, Economics and the Environment 2002). Likewise, there are many smaller scale monitoring programs that focus on particular measurements and places. Some of these can be assembled to yield a larger scale indication of trends, as shown by a recent meta-analysis of coral reef monitoring programs in the Caribbean, which showed an 80% overall rate of loss of live coral over the past 25 years (Gardner et al. 2003). We are still a long way, however, from being easily able to scale up such regional or individual studies to monitor global-scale changes in biodiversity.

That there is currently a shortage of relevant data on temporal changes in biodiversity is well illustrated by global estimates of rates of change of the extent of biomes in recent decades. These measurements are especially relevant to the WSSD/CBD target because a wide range of important attributes—such as number of species, distribution and population size of individual species, and delivery of ecosystem services—are at least partly determined by the area of the system with which they are associated. There are recent global estimates for the rate of change in area of just four (temperate and tropical forests, mangroves, and seagrass beds) of 17 biomes listed by Costanza et al. (1997). There are good regional estimates of rates of change in area of some other biomes, but there are no published global-level estimates of rate of change of intact area for freshwater swamps, lakes, rivers, estuaries, continental shelf, or coral reefs, all of which are of great importance in the provision of ecosystem goods and services (Balmford et al. 2003; Jenkins et al. 2003). Even the global estimates of recent rates of change of area for the best-studied biome in this regard, tropical forests, are subject to considerable uncertainty (FAO 2001; Matthews 2001; Stokstad 2001; Achard et al. 2002; DeFries et al. 2002). There is even less reliably quantified information on recent trends in the provision by ecosystems of services such as flood protection, nutrient cycling, carbon sequestration, and water supply. Measurements of the economic value of ecosystems and their individual components usually only consider a few of the relevant goods and services, and little is known about the changes in service delivery that occur when natural or seminatural systems are converted for human use (Balmford et al. 2002; Turner et al. 2003).

Why do we know so little that is relevant to monitoring progress toward the challenge made at the WSSD? We believe the main reasons are because existing programs to measure changes in biodiversity (1) have been disproportionately conducted in the developed world, whereas most species are found in the developing world; (2) have been designed around varied national, regional, or sectoral objectives that make them difficult to combine into global estimates; (3) have not been designed to measure or relate to a wide range of ecosystem goods and services; and (4) are opportunistic and make use of existing data on distribution and abundance collected for some other purpose. There is an urgent need to ensure that gaps in monitoring biodiversity are addressed by new programs and that both new and existing programs are suitable for their purpose.

We propose a framework that can be used to this end by scientists, providers of funding, and managers. The framework does no more than make explicit the best scientific practice to ensure that measures of biodiversity are appropriate to the purpose to which they are being applied. By measurement we refer here to some metric of the state of biodiversity. Monitoring refers to the repetition of measurements over time, and evaluation refers to the subsequent interpretation of the results.

Structure of the Framework

Our framework is a set of linked activities carried out by scientists involved in measuring and evaluating some aspect of biodiversity or an associated ecosystem function (Fig. 1). It is intended for scientists, interested parties, and stakeholders who use and are affected by the information generated and consists of three stages: (1) scoping, (2) design, and (3) implementation and reporting. A rational approach to any of the linked activities in the framework depends on the outcome of at least one other activity and, in practice, an activity may have to be repeated if feedback from other activities indicates that changes are needed.

Scoping

A major objective of biodiversity measurement is to inform policies that seek to reduce the rate of biodiversity loss as a component of a more sustainable economic development strategy. This high-level, policy-driven objective has to be translated, though, into practical activities via a decision-support mechanism. The first step in this is a scoping procedure that places the biodiversity component of the system within relevant social, political, economic, and scientific knowledge contexts. A tried and tested approach is the driver-pressure-state-impact-response (DPSIR) framework (Turner et al. 1998). This framework recognizes that at the root of environmental change are socioeconomic drivers such as exploitation of wild natural resources, agricultural intensification, urbanization, and tourism development. The cumulative effect of these, together with climate change, is pressure on environmental systems and consequent changes in the state of the environment, including biodiversity. These biological, geochemical, and physical changes will in turn have impacts (positive and negative) on human welfare. These changes in welfare affect different, often competing, interest groups in different ways and stimulate debates over equity and other ethical concerns. These in turn spur political systems into action to provide legal and management regimes to control driving forces and pressures, thus creating a dynamic cycle with feedback loops.

IDENTIFYING STAKEHOLDERS AND INTERESTED PARTIES

Stakeholders are people affected by the biodiversity of a given region. They are likely to be composed of different groups with health, welfare, intellectual, recreational, spiritual, and financial interests that will be affected by changes in biodiversity. We define *interested parties* as a subset of stakeholders who specifically support the scientific measurement of biodiversity. They may be individuals but will often be organized into associations or institutions. There may be additional stakeholders whose interests will be affected by the outcome of the evaluation but who are not among the initial set of interested parties. This may be because they do not wish to encourage a scientific study of the system, do not know about or understand it, or lack the resources or political power to influence the evaluation. Examples may include poor people, people affected by changes in ecosystem services

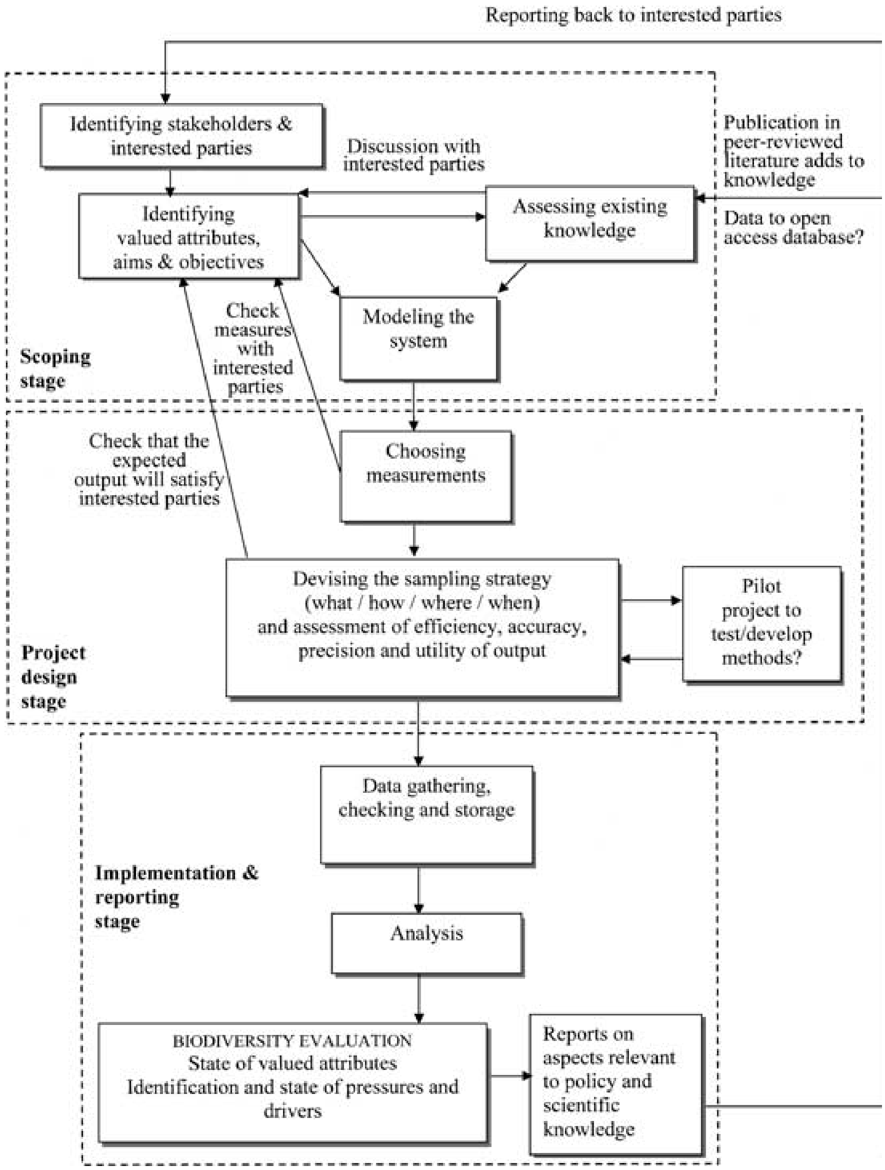

Figure 1. Framework for biodiversity evaluation. See text for details of each box.

in areas distant from where they live, and people whose options to benefit lie in the future. Scientists who measure biodiversity should attempt to identify such groups and recruit them into the group of interested parties. Even if this is not practical, they should at least ensure that their viewpoints and interests are identified and taken into consideration.

IDENTIFYING VALUED ATTRIBUTES AND OBJECTIVES

The evaluator should first decide, in consultation with the interested parties, the object of the evaluation. Several parties may be interested in the same objects (Table 1), although the valued attributes and desired states of those objects may not be similar. In evaluations of objects with controversial policy implications it is essential that interested parties are clear at the outset what the evaluation can and cannot do.

Before the measurements are defined precisely, there should be consultation with interested parties, beginning with the following straightforward questions: What do you care about? What questions do you want the evaluation to answer? What will the results be used for? When are the results needed?

Discussion of these questions is likely to lead to more detailed queries such as the following:

1. Is there a need to quantify the total amount and distribution of an attribute of interest, or will a reliable estimate based on sampling be acceptable? For a threatened species, for example, a full census of a population may be needed. For an abundant species, a reliable

Table 1. Examples illustrating the meaning of the object, interested party, and valued attribute as applied to the value of biodiversity.

Object	*Interested parties*	*Valued attributes*
Diversity of life on Earth	people who like wild nature	global species richness and the abundance, range extent, and viability of species, all of which are regarded as having existence value or nonconsumptive use value
	evolutionary biologists	global species richness and the location, abundance, and range of species as resources for research
A forested river catchment	local people whose health and livelihoods depend on a reliable water supply	volume and reliability of streamflow as a determinant of water availability to people (determined by moisture collection and retention properties of forest vegetation)
Coral reef	fishing communities	habitat for fish to be exploited on the reef or elsewhere if the reef acts a nursery or source
World population of an arctic-temperate migratory goose species	conservationists	range and population size (desired state, at least maintained well above the minimum viable level)
	farmers in the winter range	range and population size (desired state, below the level that results in significant damage to crops and grass by grazing)

estimate based on extrapolation from a sample of quadrats within the range might be adequate.

2. Where the attribute of interest is difficult to measure, is it acceptable to substitute measurements of an indicator? For instance, changes in the species richness of a single taxonomic group may, in some cases, be a good proxy for changes in the species richness of a wide range of other taxa.
   Likewise, rates of landslides onto roads might be a useful indicator for overall rates of soil erosion.
3. Do we need absolute measurements of the state of a particular attribute or can we use an index? For example, do we need to actually measure the total number of adult fish or can we substitute an index that is directly proportional to it, such as catch per unit effort?
4. What level of precision is needed and is there asymmetry in the level of uncertainty that can be tolerated? For example, for a threatened species close to its minimum viable population size, the amount of uncertainty in current population size below the best estimate is more important to those interested in its conservation than uncertainty above it.
5. What is the desired state of the attribute of interest? Asking this may help define the required precision of the estimates. In establishing a desired state, care should be taken to avoid the "shifting baseline syndrome" (Pauly 1995) by considering, wherever possible, the state of the attribute not just recently but several decades or centuries ago.
6. Is the state of the valued attribute the only thing the interested parties want to know about? This should rarely be the case if the scoping process has been fully implemented. For example, in remote-sensing surveys of deforestation rates it may be efficient to survey land use in the places converted from forest as part of the same monitoring program, thereby shedding important light on drivers as well as state, at modest extra cost.

The process of identifying valued attributes, taken together with those of assessing existing knowledge and modeling the system should lead to the identification of a clear set of aims and objectives for the evaluation. These are likely to be altered and refined during later steps in the framework as practical constraints become apparent. However, all but the most trivial of such changes should be discussed with interested parties.

Assessing Existing Knowledge

Existing knowledge may come from previous scientific studies of the system, or it may take the form of assumptions, based on studies of other systems, if the similarities are sufficiently strong. Even if the information is derived from the same system, it may become an unreliable guide because of changes over time. For example, the sustainability of a bushmeat harvest from a tropical forest may be evaluated by linking data on current harvest levels with preexisting knowledge of the geographical distribution of hunting areas. If, unknown to the evaluator, hunting areas shift, genuine overharvesting may be masked, leading to unreliable conclusions about the sustainability of the offtake.

Existing knowledge can be used to make sample surveys more efficient. For example, to measure changes in the threat status of species, prior knowledge of geographical range and previous assessments of threat can be used as the basis for stratified random sampling for survey plots.

The initial stages of an evaluation of biodiversity may (and often should) involve alternating periods of examination of existing knowledge and discussion with interested parties of the conclusions drawn from it. Existing knowledge is likely to be challenged by interested parties during this process, which therefore helps refine the objectives of the proposed work and identify parts of it in which faulty assumptions or lack of precision are crucial.

Modeling the System

Consideration of existing knowledge and the requirements of interested parties provides a qualitative background for the design of the evaluation, but there is an important intermediate step: the preparation of a model of the system. This may seem unnecessary and premature, especially if existing information is sparse. It might seem better to collect data first and then use it to build a model. We think this view is mistaken because the development of even a simple conceptual model forces assumptions to be recognized and made explicit and highlights defects in the reasoning linking the attributes of interest to the measurements that can be made. Indeed, the decisions scientists make about how to conduct a study are always guided by some sort of model of the system. At issue is only whether the model is explicit and accessible to stakeholders or is a hidden web of undeclared assumptions.

The nature of an appropriate model depends on the system under consideration and the attributes to be measured. Suppose the object of the evaluation is a population of animals or plants and the attributes to be measured are population size and trends over time. An appropriate model might then be a matrix model of the numbers of individuals in various life-history stages over a period of several years. Changes in numbers are modeled as outcomes of time and stage-specific survival and reproductive rates (and immigration and emigration if the population is not closed).

If the evaluation aims to measure key pressures and drivers, it is important that the model of the system include the relationships among these, ideally within a driver-pressure-state-impact-response framework, even if it is only possible to do so in conceptual form.

Design

Choosing Measurements

With a model in place, the evaluator can consider what specifically to measure by extending the model to include relationships between valued attributes and things that can actually be measured. In the case of a bird population, for instance, the total population size may best represent the state of the valued attribute, but it may only be possible to obtain reliable counts for singing males during the breeding season. The evaluator either assumes a census of singing males provides an acceptable index of the population size or initiates a study to calibrate this index with respect to the size of the whole population.

Researchers have a vast range of potential measurements to choose from, ranging from those that summarize general characteristics of biomes to those that examine changes in distribution, population size, and gene frequencies of a single species. Detailed evaluation of the strengths and weakness of different measurements is beyond the scope of this paper, but the decision should be made with reference to two key factors: (1) fitness for purpose, which comprises both the relevance to interested parties of what is measured, as identified by the scoping stage, and the adequacy of the measurement in terms of the precision and accuracy of estimates and (2) availability of expertise, equipment, and money, which we call available effort. We illustrate a procedure to evaluate these two factors with a hypothetical example.

To estimate the area of old-growth native forest within a large region, we could either send teams of surveyors to classify and map hundreds of sites on the ground directly or use satellite imagery. Suppose satellite-based surveys are cheaper but, even with ground truthing, the results are less reliable because native forests are difficult to distinguish from plantations of exotic species. Step 1 of the procedure is to plot the expected fitness for purpose of the results against the completeness of coverage of each of the alternative survey programs (Fig. 2a). Although both methods become more accurate as they approach complete coverage, the satellite-based estimates are only half as useful as the direct surveys, even when complete, because of their lower reliability.

Step 2 is to consider the completeness of the survey program in relation to available effort (Fig. 2b). Here, the satellite-based method wins because a given level of coverage requires much less effort than the direct ground-based survey. The third step is to combine steps 1 and 2 to estimate the relationship between fitness for purpose and available effort (Fig. 2c). Suppose we have 20 units of effort and neither survey program has yet been started. Figure 2c shows that the satellite-based survey (which could be completed with this much effort) is more useful than the incomplete direct survey, in spite of the higher reliability of direct survey data. However, had 40 units of effort been available the usefulness of the results of the direct survey would have been higher. Intuition alone would not necessarily lead to this conclusion because the answer depends on the particular functions that relate fitness for purpose to completeness and completeness to effort. Different outcomes can be obtained with different functions.

A final consideration in choosing among alternative methods of measurement is the amount of information that is already available. For example, suppose 20 units of effort had already been expended on the direct survey before the satellite-based method became available. As a result the ground survey is about 5% complete. How should we spend further resources that become available? Continuing to spend resources on the direct survey will yield

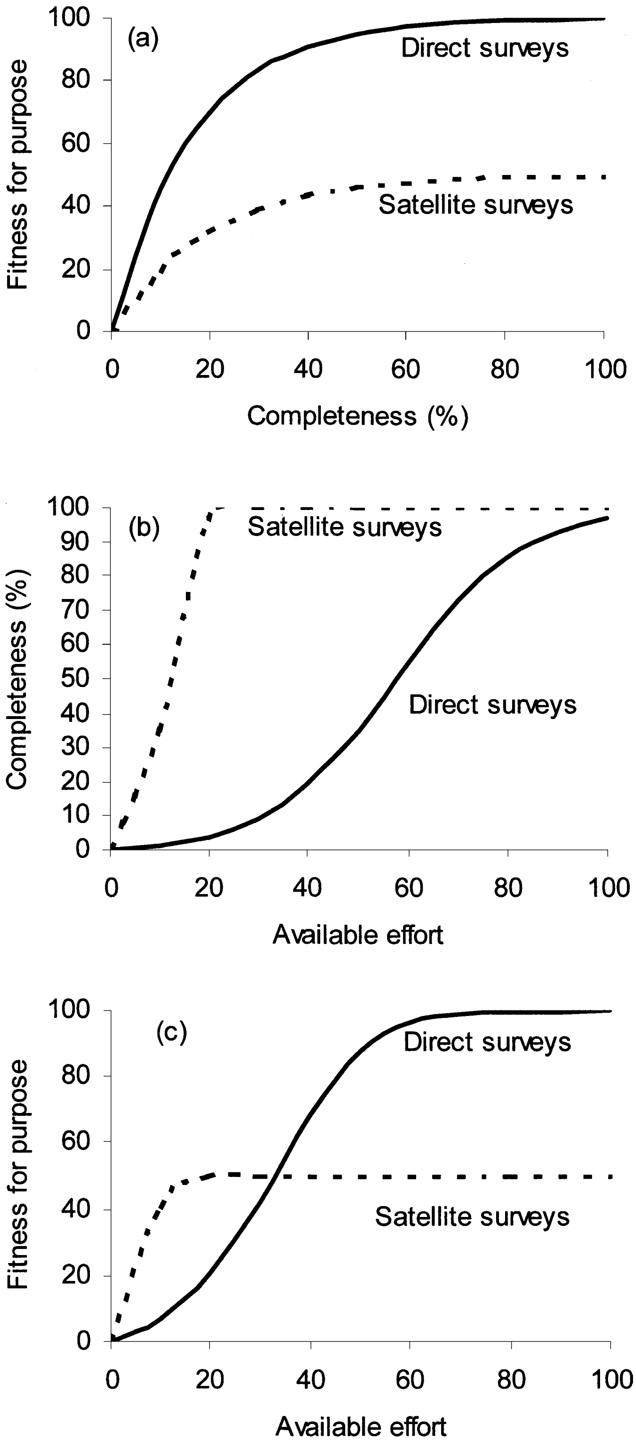

Figure 2. Hypothetical relationships of (a) the fitness for purpose of two different approaches to surveying forest cover versus completeness, (b) relationship of completeness to available effort, and (c) relationship of fitness for purpose to available effort. Fitness for purpose and available effort are scaled arbitrarily from 0 (minimum) to 100 units (maximum).

a more useful result than switching to the satellite-based method provided that at least 10 more units of effort are available (Fig. 2c). However, if <10 units are available it might be better to switch to the satellite-based survey.

This example illustrates the value of having some prior sense of the relationships between the variables in Fig. 2 and of considering the total amount of time and resources available and the results that have already been obtained by previous efforts. If repeat surveys to monitor change are anticipated, then the ease with which a given evaluation could be exactly reproduced in future would need to be incorporated into its fitness for purpose versus completeness function. In our example, this might reduce the advantage of the direct ground surveys (for which it might be difficult to train new staff to reproduce the forest classification method exactly) over the more automated processing of satellite imagery.

Sampling Strategy

Having decided the type of measurement to make, the next task is to use the model to develop a sampling strategy that specifies exactly what, when, where, and how to measure. Is a complete survey needed, or will a sample suffice? In the latter case, the precision of resulting estimates can be improved by dividing the potential sampling units into strata based on prior knowledge of spatial variation in the attribute being surveyed. The proportion of the area within each stratum that is selected for survey can then be chosen to maximize the precision of the estimate. Such stratified random sampling might involve disproportionate sampling in areas known to support high population densities (in the case of a population survey), high concentrations of endemic species (in a species survey), high concentrations of crop pollinators (in a survey of ecosystem services), or high rates of land-use change (in a survey of change of biome area). Stratification may be worthwhile even if some of the prior information on which it was based was faulty. The resulting estimates are still likely to be more precise than if stratification was not attempted, though less precise than if the prior information had been perfect.

Because one will nearly always be interested in trends over time, careful consideration must be given to the extent to which attributes fluctuate over short time intervals and to showing systematic trends over longer periods. The model's assumptions about the causes of variation over time in the attributes being measured are important here. For example, trends in the population size of elephants, with their high mean annual survival rate and low average fecundity, would probably be well described by accurate surveys at intervals of several years. However, the population sizes of voles, with low annual survival and high fecundity, fluctuate greatly, so annual surveys would be needed to discern a reliable trend over the same time period. Because life-history and demographic traits can be predicted to some extent from body size and ecological variables, an assessment of the appropriate frequency of estimates can be made even when data on the species of interest are sparse.

The model of the system can be used to simulate the likely outcomes of the evaluation, after making some plausible guesses about the actual state of the attributes being measured. Does it still appear the monitoring program will yield meaningful results that will help answer the questions posed by the interested parties? Development work on the data collection methods or a pilot survey to check that they work as envisaged might be needed.

Implementation and Reporting

Data Gathering

Nine general principles are important to the practice of biodiversity data collection for a wide range of applications:

1. Ensure that people collecting data are adequately trained and follow a common written protocol for collecting and recording information, which does not change over time.
2. Keep raw data for checking and reinterpretation.
3. Store data in their rawest form, rather than just as worked-up totals or averages.
4. Georeference field study areas.
5. Record sampling effort and who collected the data.
6. Record both presence and absence in surveys of distribution and abundance.
7. Carry out checks to ensure errors in recording and data storage are kept to an acceptable level.
8. Where possible, collect additional, low-cost data that may be useful later.
9. Review progress regularly to check that the data being collected will address the questions originally posed.

Analysis, Evaluation, and Reporting

As with data gathering, the details of the analysis and reporting of the evaluation will be specific to each particular case, but we again make some recommendations on issues of wide significance.

1. The sensitivity of the conclusions of the evaluation to the underlying model and its assumptions should be explored and reported clearly. Where appropriate, alternative conclusions arrived at from plausible variants of the model should be reported and the differences in model structure or assumptions that give rise to them identified.
2. The results of the study should be used to update the underlying model of the system as a basis for future evaluations. Defects in the model should be identified clearly and remedies suggested.
3. The survey design, procedures used in sample area selection, and the fieldwork and analysis protocols should all be described in sufficient detail that the survey can be repeated. This is vital if stakeholders are interested in trends over time and is especially important for complex, semiautomated techniques such as the mapping of biomes.
4. Where possible, the raw data from the survey should be available in paper or electronic form to other stakeholders and evaluators for alternative analyses.
5. Precise survey localities should be archived as geographical coordinates and maps.
6. The results of the survey should, wherever possible, be published in the peer-reviewed literature. Where this is not possible, an attempt should be made to subject the outputs to other forms of external review.

Reporting to Interested Parties

Those making scientific evaluations of biodiversity should report the results to interested parties in ways that minimize misinterpretation. Reports should address specifically the possible policy responses under consideration, including the relative strength of the evidence for and against each option and the extent to which this could be altered by violations of assumptions.

Reporting of the results of a survey to interested parties will often lead to proposals for a repetition of the evaluation or for some connected piece of research. Within the framework, this should involve returning to the beginning and discussing the aims and objectives with interested parties again. The evaluation just completed now becomes part of the existing knowledge to be taken into account when designing the new study. The driver-pressure-state-impact-response framework may be modified and new stakeholders identified and recruited as interested parties. With the new information it will usually be possible to identify important new questions and identify changes to methods that will improve efficiency or accuracy. Such changes in objectives and methods should be considered carefully and with great caution. Because measurement of changes over time is fundamental to monitoring it will usually be essential to ensure that the original design is repeated to ensure that results can be compared directly among successive evaluations. If there is a strong case for using new methods, then the results must be calibrated so that they can be made comparable to previous studies.

Conclusions

This framework highlights the fundamental importance in biodiversity measurement of ensuring that the data gathered and the analyses done on the data answer the questions relevant to interested parties with the required accuracy, precision, and spatial and temporal scope. In our view, the extent of global data gathering underway is inadequate to meet the challenge set out at the WSSD in Johannesburg. Although excellent monitoring programs exist, the results they provide are simply too patchy and unrepresentative in terms of the taxa, biomes, parts of the world, and aspects of biodiversity that they measure for them to provide policy makers with a clear, broad, and reliable picture of people's changing impacts on wild nature and the consequences for human welfare (Balmford et al. 2003; Jenkins et al. 2003). There are many reasons for the dearth of appropriate monitoring systems, but we believe insufficient attention to the questions most relevant to interested parties and stakeholders lies at the root of the problem. The need to connect with stakeholders in a more explicit way is also becoming more widely recognized in systematic conservation planning (Cowling & Pressey 2003).

We therefore believe that as a first step the community of interested parties, including intergovernmental, governmental, and nongovernmental organizations and academic conservation scientists, must undertake a rapid review of existing programs of biodiversity monitoring at the global level. This will require the synthesis of otherwise scattered data, which will in turn make them more readily available and more useful. Existing programs of data collection and analysis can then be evaluated against the needs of key stakeholders, and in particular, against the needs of signatories to the World Summit on Sustainable Development and parties to the CBD. This evaluation should in turn yield two important outcomes.

First, the review should identify existing monitoring programs that are already contributing to or delivering robust, global measures of the changing state of wild nature, such as the forest monitoring programs of the U.N. Food and Agriculture Organization and the European Union's Joint Research Center (FAO 2001; Achard et al. 2002). Where they have shortcomings, the review should identify them and propose remedies. These programs must continue and where possible their coverage should be enhanced. Information about their strong points should be disseminated widely to promulgate best practice.

Second, this synthesis should identify key gaps in both monitoring and scientific understanding. Priorities should include poorly monitored aspects of biodiversity that are central to the Summit's goals, such as the provision of ecosystem services (such as regular access to clean water and storm protection), and gaps in more traditional monitoring of those taxa and biomes that are particularly important in terms of their contribution to global biodiversity or human welfare. These should be addressed by the development of realistic new programs capable of delivering substantial improvements in knowledge of otherwise poorly understood geographic areas, biomes, groups of organisms, and aspects of biodiversity. Such programs must be implemented urgently and with realistic goals for completion. There should be a particular focus on establishing repeatable estimates of rates of change against which progress toward reducing rates of biodiversity loss by 2010 can be measured.

New monitoring programs and improved analysis and expansion of existing ones will require a marked increase in funding, capacity, and cooperation among nongovernmental organizations, academics, and governmental and intergovernmental agencies. Constraints on resources reinforce the potential value of the framework we developed for improving the fitness for purpose and value for money of biodiversity monitoring. In turn, increasing the efficiency of monitoring is key to tracking progress against the 2010 target, to quantifying how far our conservation efforts are working, and, ultimately, to slowing the rate of biodiversity loss.

Acknowledgments

This paper arose from work carried out by the Royal Society Working Group on Measuring Biodiversity for Conservation. We are grateful to all members of the group for useful discussions and to the Royal Society for initiating and supporting the group's activities.

Literature Cited

Achard, F., H. D. Eva, H. -J. Stibig, P. Mayaux, J. Gallego, T. Richards, and J. -P. Malingreau. 2002. Determination of deforestation rates of the world's humid tropical forests. Science **297:**999–1002.

Balmford, A., et al. 2002. Economic reasons for conserving wild nature. Science **297:**950–953.

Balmford, A., R. E. Green, and M. Jenkins. 2003. Measuring the changing state of nature. Trends in Ecology & Evolution **18:**326–330.

Cowling, R. M., and R. L. Pressey. 2003. Introduction to systematic conservation planning in the Cape Floristic Region. Biological Conservation **112:**1–13.

Costanza, R., et al. 1997. The value of the world's ecosystem services and natural capital. Nature **387:**253–260.

Daily, G. C. 1997. Nature's services. Societal dependence on natural ecosystems. Island Press, Washington, D.C.

DeFries, R. S., R. A. Houghton, M. C. Hansen, C. B. Field, D. Skole, and J. Townshend. 2002. Carbon emissions from tropical deforestation and regrowth based on satellite observations for the 1980s and 1990s. Proceedings of the National Academy of Sciences of the United States of America **99:**14256–14261.

Food and Agriculture Organisation (FAO). 2001. Global forest resources assessment 2000: main report. FAO, Rome.

Gardner, T. A., I. M. Côté, J. A. Gill, A. Grant, and A. R. Watkinson. 2003. Long-term region-wide declines in Caribbean corals. Science **301:**958–960.

Jenkins, M., R. E. Green, and J. Madden. 2003. The challenge of measuring global change in wild nature: are things getting better or worse? Conservation Biology **17:**1–4.

Matthews, E. 2001. Understanding the FRA 2001. World Resources Institute, Washington, D.C.

Millenium Ecosystem Assessment. 2004. Available from http://www.

millenniumassessment.org/en/index.aspx (accessed 21 September 2004).
Pauly, D. 1995. Anecdotes and the shifting baseline syndrome of fisheries. Trends in Ecology & Evolution **10:**430.
Reid, W. V., and G. M. Mace. 2003. Taking conservation biology to new levels in environmental decision-making. Conservation Biology **17:**943–945.
Secretariat of the Convention on Biological Diversity. 2003. Handbook of the convention on biological diversity. 2nd edition. Secretariat of the Convention on Biological Diversity, Montréal.
Stokstad, E. 2001. U.N. report suggests slowed forest losses. Science **291:**2294.
The H. John Heinz III Center for Science, Economics and the Environment. 2002. The state of the nation's ecosystems. Cambridge University Press, Cambridge, United Kingdom.
Turner R. K., I. Lorenzoni, N. Beaumont, I. J. Bateman, I. H. Langford, and A. L. McDonald 1998. Coastal management for sustainable development: analyzing environmental and socio-economic changes on the UK coast. The Geographical Journal **164:**269–281.
Turner, K., Paavola, J., Farber, S., Cooper, P., Jessamy, V., Rosendo, S. and Georgiou, S. 2003, Valuing nature: lessons learned and future research directions. Ecological Economics, **46:**493–510.
United Nations. 2002. Report of the World Summit on Sustainable Development, Johannesburg, South Africa.