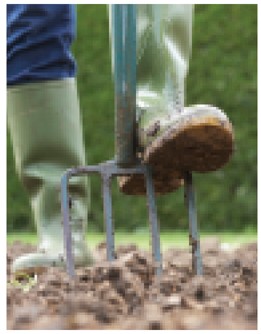
MEASURING PROGRESS

A Practical Guide from the Developers of the Environmental Performance Index (EPI)

Yale Center for Environmental Law & Policy



Acknowledgments

Authors

Angel Hsu
Director,
Environmental Performance Measurement Project
Yale Center for Environmental Law & Policy (YCELP)

Laura Johnson
Research Associate
Yale Center for Environmental Law & Policy (YCELP)

Ainsley Lloyd
Research Associate
Yale Center for Environmental Law & Policy (YCELP)

Contributors

Alex de Sherbinin
Senior Research Associate
Center for International Earth Science Information Network (CIESIN), The Earth Institute,
Columbia University

William Dornbos
Director
Environment Northeast - Connecticut

Jay Emerson
Associate Professor
Department of Statistics,
Yale University

Daniel C. Esty
Hillhouse Professor of Environmental Law & Policy
Yale University

Nora Hawkins
Research Associate
Yale Center for Environmental Law & Policy (YCELP)

Marc Levy
Deputy Director
Center for International Earth Science Information Network (CIESIN), The Earth Institute,
Columbia University

Reviewers

Alex de Sherbinin
Senior Research Associate
Center for International Earth Science Information Network (CIESIN), The Earth Institute,
Columbia University

Jay Emerson
Associate Professor
Department of Statistics,
Yale University

Josh Galperin
Associate Director
Yale Center for Environmental Law & Policy (YCELP)

Aaron Reuben
Research Associate
Yale Center for Environmental Law & Policy (YCELP)

Fiona Schwab
Senior Advisor
Office of Strategic Affairs (Abu Dhabi, United Arab Emirates)

Tanja Srebotnjak
Senior Fellow
Ecologic Institute

Susanne Stahl
Program Coordinator
Yale Center for Environmental Law & Policy (YCELP)

July, 2013

Hsu, A., L.A. Johnson, and A. Lloyd. 2013. *Measuring Progress: A Practical Guide From the Developers of the Environmental Performance Index (EPI).* New Haven: Yale Center for Environmental Law & Policy.

Measuring Progress:
A Practical Guide From the Developers of the Environmental Performance Index (EPI)

Table of Contents

1 Introduction

Introduction

We are currently in an age of information dilemmas. Although data and statistics are now more accessible than ever due to technological advances, we still suffer from knowledge gaps and information asymmetries. While these gaps will always persist, the challenge lies in bridging scientific understanding with needed data collection to effectively manage an issue at multiple scales–locally, regionally, and globally. Particularly with respect to the environment, decisions have historically relied too heavily on "educated guesses but not hard facts," allowing critics to dismiss the importance of pollution control and natural resource management.[1]

Environmental **indicators** and performance indices are emerging as powerful tools for decisionmakers to navigate the uncertain information landscape. They distill complex information, allowing decisionmakers and key audiences to efficiently spot critical areas of concern, support policy development and target setting, and measure impacts of policy responses.

Over the past two decades, there have been dramatic increases in the development and use of indicators as cost effective and efficient means to inform decisionmaking and management.[2,3] Gross Domestic Product (GDP), created in 1934, has become a ubiquitous measure of economic progress. The Human Development Index (HDI) and Millennium Development Goals (MDGs), among many others, now serve as similar initiatives to GDP to help gauge progress toward social and economic development.

This manual serves as a guide to the process of developing an **environmental performance** index based on the experience of the Yale Center for Environmental Law and Policy (YCELP) and the Center for International Earth Science Information Network (CIESIN) at Columbia University in developing the Environmental Sustainability Index (ESI) and its subsequent form, the Environmental Performance Index (EPI). Here we share our "lessons learned" in creating performance indices, particularly with respect to environmental issues.

We have learned many of these lessons through our own collaboration and analyses, as well as extensive consultation with subject-area experts, decisionmakers, journalists, and researchers. Although we admit that the EPI is an ever-evolving process, we believe we have valuable insights to share. We also hope that this manual will start to cultivate a conversation about the methods and challenges of developing and maintaining indicators and indices.

Why the EPI?

The EPI ranks countries on performance indicators tracked across policy categories that include both environmental public health and ecosystem vitality. These indicators, aimed at the national government scale, provide a gauge of a country's performance toward policy goals.

Although the EPI is primarily applied at the country-level, its flexible framework and methodology can also be adapted for a variety of purposes at different scales. Individual countries have approached the YCELP and CIESIN teams to develop sub-national environmental performance indices that rank provinces or cities. Others have sought to apply the EPI toward a particular objective (e.g., measuring and

1 Esty, D.C. 2001. Toward Data-Driven Environmentalism: The Environmental Sustainability Index. *The Environmental Law Reporter: News & Analysis.* May 2001.

2 USGCRP (US Global Change Research Program). 2012a. *Climate Change Impacts and Responses: Societal Indicators for the National Climate Assessment.* NCA Report Series, Volume 5c. Washington DC: US Government Printing Office.

3 USGCRP (US Global Change Research Program). 2012b. *Ecosystem Responses to Climate Change: Selecting Indicators and Integrating Observational Networks.* NCA Report Series, Volume 5a. Washington DC: US Government Printing Office.

comparing economic competitiveness in relation to the environment). Since the ESI's conception in 2000, there has been a proliferation of index and indicator efforts, with particular attention paid to the environment and sustainability. One review noted over 500 sustainability-related indicator efforts in existence in 2003. Of these efforts, 67 were global, 103 were national, 72 were state or provincial, and 289 were local or metropolitan.[4] These included the Dow Jones Sustainability Index, UNDP's Human Development Index, the World Wildlife Fund's Living Planet Index, and the World Economic Forum's Competitiveness Index, among others. *Box 1.1* provides an overview of some of these indices and their aims.

The EPI stands out from other quantitative environmental efforts. It incorporates a comprehensive picture of high-priority environmental issues, including resource consumption, depletion of environmental assets, pollution, species loss, and other important topics. Other mainstream environmental assessment efforts do not quite achieve this scope and are outside the purview of this manual. For instance, the Ecological Footprint (EF), a well-known effort to quantify human environmental impacts, is based on an architecture that includes natural resources related to consumption. The EF is aimed at addressing long-term environmental impacts stemming from human demands, but omits non-consumption high-priority issues, such as pollution and waste management. While some green accounting methods, such as Green GDP, are internationally accepted statistical standards, many remain controversial in terms of consistency in methodological calculation and results.[5] In addition, some environmental accounting efforts are based on environmental assets that are commercially exploited and quantify impacts in economic terms, resulting in highly debatable and often controversial results that limit widespread policy adoption.

The strength of the EPI is in its expert consensus-based framework that identifies critical environmental policy issues and calculates scientifically rigorous metrics on a common and comparable scale. The framework is flexible enough to include almost any issue deemed a high priority. Because the metrics in the EPI rely on a country's performance relative to a defined policy target, it provides a basis for comparisons regardless of the data or issue at hand.

The strength of the EPI is in its expert consensus-based framework that identifies critical environmental policy issues and calculates scientifically rigorous metrics on a common and comparable scale. The framework is flexible enough to include almost any issue deemed a high priority. Because the metrics in the EPI rely on a country's performance relative to a defined policy target, it provides a basis for comparisons regardless of the data or issue at hand.

The EPI's methodology and principles are based on developing performance indicators calculated as a country's distance to a defined target. These characteristics have made the EPI an oft-cited reference for policymakers, the media, and members of the research community.[6] Our success has been due in no small part to the contributions of hundreds of experts who have convened to develop and vet our process, data and methodology. Moreover, experience in refining the EPI over the past 13 years has provided the members of YCELP and CIESIN with valuable expertise in the realm of environmental performance measurement and composite index development.

---

4 Parris, T.M. and R.W. Kates. 2003. Characterizing and Measuring Sustainable Development. *Annual Review of Environmental Resources.* 28:559–86.

5 Li, L. 2007. China postpones release of report on 'Green' GDP accounting. Worldwatch Institute. www.caep.org.cn/english/ReadNewsEN.asp?NewsID=1069. Last accessed: January 28, 2013.

6 See www.epi.yale.edu/community/news for a full listing of news articles and reports that cite or reference the EPI.

Box 1.1. Examples of Environmental Indices

OECD Better Life Index
www.oecdbetterlifeindex.org
Measures "well-being" across eleven categories, including housing, income and jobs for all thirty-four OECD countries.

Ocean Health Index
www.oceanhealthindex.org
Evaluates the condition of marine ecosystems according to ten human goals, which represent the key ecological, social, and economic benefits that a healthy ocean provides.

Ecological Footprint
www.footprintnetwork.org
A metric that represents the amount of biologically productive land and sea area necessary to supply the resources a human population consumes, and to assimilate associated waste.

Human Development Index
www.hdr.undp.org/en/statistics/hdi
Measures human development in three basic dimensions: long and healthy life (health), access to knowledge (education), and a decent standard of living (income).

WEF's Global Competitiveness Index
www.weforum.org/issues/global-competitiveness/index.html
Gauges 144 economies according to a set of institutions, policies, and factors that determine levels of productivity.

History and Evolution of the EPI

YCELP, in partnership with CIESIN and the World Economic Forum, has been developing global indices of environmental sustainability and performance for more than a decade. When we first published the ESI (the predecessor of the EPI) in 2000, governments around the world were struggling to mainstream sustainability into their policy goals. The Millennium Declaration promulgated earlier that year by the United Nations sought to articulate a comprehensive vision for improving quality of life in developing countries. This vision involved long-term development goals that were grounded in quantitative metrics to be achieved by 2015. While the Declaration included environmental sustainability as a goal, it contained virtually no relevant quantitative metrics to support the goal. This stood out in sharp contrast to the other goals of the Declaration, such as poverty reduction, health care improvement, and education expansion, each of which could rely on well-established metrics for policy setting. The ESI was, in part, a response to this data gap.

The ESI was a first attempt to rank countries on multiple components of environmental sustainability, including natural resource endowments, past and present pollution levels, environmental management efforts, contributions to the protection of the global commons, and a society's capacity to improve environmental performance over time.

Because of its broad scope and unclear definition of sustainability, the original ESI proved to have limited utility as a pragmatic tool for priority setting in policy. The large number of indicators made it difficult to make targeted policy assessments and recommendations. To address this challenge, the Yale-Columbia research team developed the 2006 pilot EPI as a departure from attempting to measure sustainability, and instead aimed to measure performance. The narrower scope of the EPI focuses more closely on current policy action as opposed to historical trends and endowments, both of which comprised the ESI. It was our hope that the core set of metrics comprising the EPI would allow governments to be held accountable for sustainability impacts, as good environmental performance is a prerequisite for environmental sustainability. Through developing the EPI we found that policy drivers could be identified and quantified as outcome-oriented indicators and offer a more relevant approach than sustainability indicators, which were not gauged against clearly defined targets.

> The EPI methodology was crafted in a way that would allow countries to compare their strengths and shortcomings against their peer countries.

Additionally, the EPI methodology was crafted in a way that would allow countries to compare their strengths and shortcomings against their peer countries. The pared down EPI measured 16 indicators – rather than the unwieldy 76 of the ESI – across six identified policy categories sorted into two overarching environmental objectives: 1) reducing environmental stresses on human health and 2) promoting ecosystem vitality and sound natural resource management. These two objectives were selected because the EPI was designed to be a useful tool for policymakers, and priorities with respect to the environment are often thought of along lines of people and ecosystems. The EPI continued to evolve through later releases, and the 2008 and 2010 reports included changes in data and methodology based on expert feedback. The 2008 EPI specifically engaged a number of leading experts as consultants on indicator development for policy categories. We increased the number of indicators to 25 in the 2008 and 2010 EPIs, and we introduced imputations for missing data to increase the coverage of included countries. Another methodological change involved adjustments to the weights given to certain policy categories, with the greatest weight attributed to the climate change category. The climate change category weight was increased from one-tenth in the pilot EPI score to one-quarter of the 2008 and 2010 EPI scores because of the prominence of climate change as a global issue and its overarching influence over other ecosystem policy issues. In the 2010 EPI we modified our **aggregation** methodology to include **logarithmic transformation** of certain indicators that effectually "spread out" data to highlight more subtle distinctions among leading countries.

By and large, the framework and approach to performance measurement remained relatively consistent from 2006 to 2010, with one major shortcoming: the inability to produce results that may be comparable over time. Progressive versions of the EPI provided only a snapshot of environmental performance at a given time because of alterations in data and methodology. To address this shortcoming we changed indicator criteria for the 2012 EPI, most notably requiring the availability of time series data for an indicator to be included. The new criteria focuses on a smaller set of core indicators that meet standards such as direct measurement (rather than modeled data), consistent time series, and institutional commitments to maintain data streams into the future. The application of these more stringent criteria enabled us to track performance over time and provide a foundation to continue tracking performance using a more consistent set of indicators into the future.

These changes to the EPI's methodology also allowed us to develop the Pilot Trend Environmental Performance Index (Trend EPI). The Trend EPI ranks countries on the change in their environmental performance over the last decade. As a complement to the EPI, the Trend EPI shows which countries are improving and which countries are declining over time, in terms of performance. By using the Trend EPI, countries are able to better assess their environmental progress through time and determine the efficacy of policies implemented to address issues surrounding their performance.

These changes to the EPI's methodology also allowed us to develop the Pilot Trend Environmental Performance Index (Trend EPI). The Trend EPI ranks countries on the change in their environmental performance over the last decade. As a complement to the EPI, the Trend EPI shows which countries are improving and which countries are declining over time, in terms of performance. By using the Trend EPI, countries are able to better assess their environmental progress through time and determine the efficacy of policies implemented to address issues surrounding their performance.

Who Should use this Guide?

This guide is based on the framework of the EPI and is meant for those who are specifically seeking guidance on how to develop environmental performance indicators or aggregated, composite indices. This guide is not intended for an audience generally interested in understanding the use of indicators or as a widespread survey of indicator calculation methodologies. We also do not intend to duplicate any of the material found in other detailed handbooks on statistical indicator construction, such as those from the OECD (see *Appendix 1* for a listing of other useful resources on indicator construction). Instead, the motivation for this manual is to reflect upon more than a decade's worth of our experience in constructing and refining the EPI.

Structure of this Guide

This manual is organized as follows:

Chapter 2: **'Establishing the Basis of an Index'** discusses how to shape the foundation of an index, which includes identification of the goals, target audience, and guiding criteria for an environmental performance index.

Chapter 3: **'Preparing the Process'** describes how to create a working team to manage the design process for an index, as well as how to engage experts and other stakeholders.

Chapter 4: **'Building an Indicator Framework'** provides step-by-step methods for defining an indicator framework, including what perspectives and criteria to consider.

Chapter 5: **'Evaluating Data Quality'** includes strategies for obtaining high quality data to create a credible index.

Chapter 6: **'Calculating an Index'** discusses the step-by-step details of index construction from indicator calculation and weighting selection to final aggregation.

Throughout each of these chapters we provide insights from the development of the EPI, including details on how we analyzed our own options and what criteria we used to design the EPI. *Figure 1.1* provides a generalized diagram of this process.

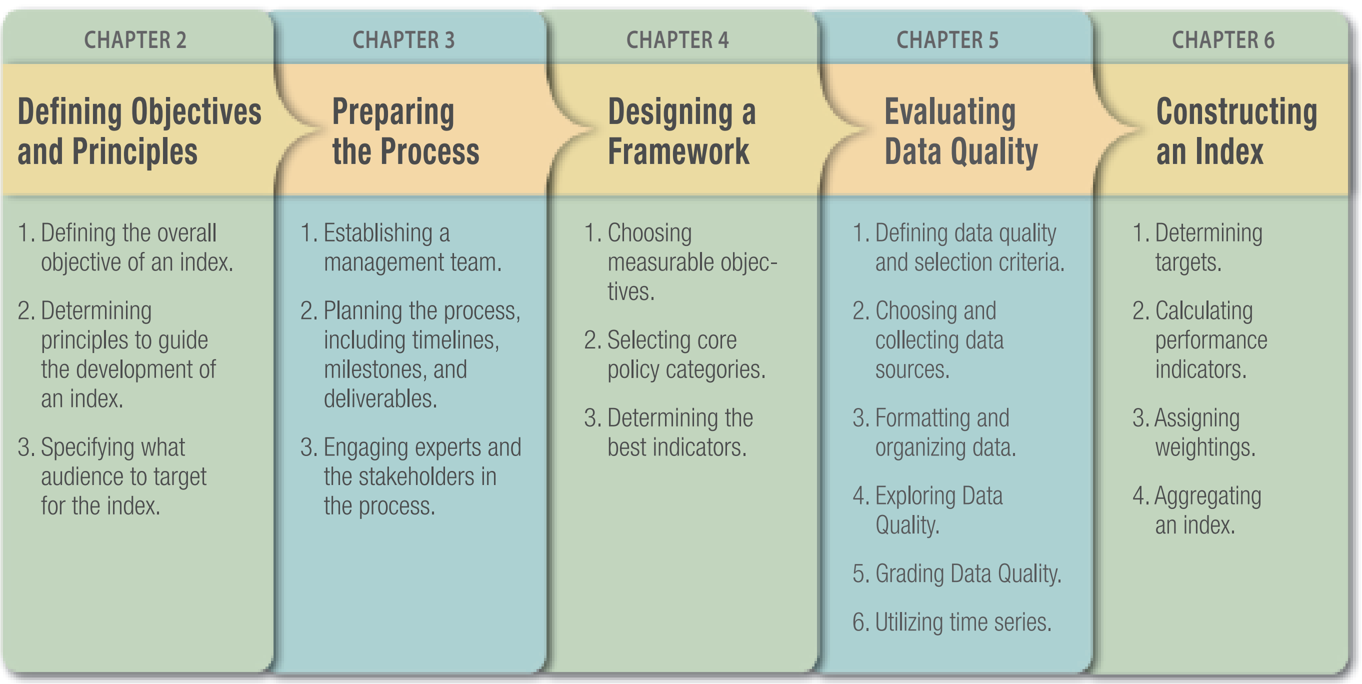


Figure 1.1. An organizational roadmap describing the major steps of the EPI design process, as described in this manual.

Box 1.2. Key Definitions

We define **"environmental performance"**as measurable results of an entity's management of environmental issues.

We use **"environmental indicator"** to describe a quantitative variable measured or calculated from data that is used to identify pressures on the environment, environmental conditions (states), or policy responses, or trends thereof.

Combining the above two definitions, **"environmental performance indicators"** refer to applying environmental stress, state, and policy response against clearly defined targets. The definitions of targets and steps for setting targets are discussed in *Chapter 6: Constructing an Index*.

We use **"index"** as an aggregate of performance indicators, which generally implies conversion to common units (or a unitless scale) and weighted or unweighted aggregation (i.e., averaging, adding, or applying other mathematical operations). A collection of indicators then specifies an architecture that identifies high-priority issues with all metrics calculated on a common scale.

Other keywords, denoted in bold, are defined in the Glossary at the end of this guide.

2 Establishing the Basis of an Index

Topics covered in this chapter

- **Overall Goals and Aims**
  Defining the purpose and message of an index.
- **Audience**
  Identifying the target audience for which an index is aimed.
- **Principles**
  Determining underlying principles that will underpin the foundation of an index.

Introduction

This chapter seeks to lay a critical foundation for what an **index** is ultimately trying to measure, who it is intended to reach, and what underlying **principles** can guide the subsequent development of an index. In theory this process should be iterative, as illustrated in *Figure 2.1*, with the consideration of goals, audience, and principles informing the process of developing an index as it evolves.

Overall Goals of an Index

The first challenge in developing an environmental performance index is to define its **goal** or purpose in conceptual terms. What is the index trying to achieve? The answer to this critical question will facilitate the definition of an index's overall aims, which provide an essential foundation for other components of the index, including the design of the indicator **framework** and data selection.

Figure 2.1. The relationships between an index's goals, principles, and audience and how each informs the process of developing an index.

The following are examples of common purposes and benefits for which environmental performance indices are created:

▶ Measurement

- Measure or gauge something.
  - Determine the effectiveness of an air pollution control policy.
- Track progress toward identified goals or targets.
  - How close a city is to achieving greenhouse gas reductions.

▶ Communication

- Communicate information to a specific audience.
  - Policymakers, scientists, or the public.
- Translate complex information for social or policy learning.
  - Simplify complicated data to the lay public.
- Call attention to the most relevant or current policy-related issues.
  - High-priority topics related to vulnerable populations in a climate change adaptation index.

▶ Comparison

- Develop a common **metric** by which to compare progress or performance between entities.
  - An index that **normalizes** pollutant concentrations to compare air quality between cities.
- Determine impact of policies or processes.
  - Compare performance before and after the implementation of a standard or regulation.

Box 2.1. Developing an Index to Measure Gender Equality and the Environment: IUCN's Environmental Gender Index

Several comprehensive international agreements, including all three of the Rio Conventions and the Convention on the Elimination of all Forms of Discrimination Against Women (CEDAW), include mandates on gender and the environment. However, many governments are struggling to integrate a gender approach into environmental decisionmaking. To help overcome these challenges, the IUCN Global Gender Office recently initiated the Environmental Gender Index (EGI) as a new tool to help assess and measure gender equality and women's empowerment in the context of global environmental agreements.

The EGI is geared toward becoming an important mechanism for policymakers and governments. The aim of this index, "to promote greater transparency and accountability, and to contribute to the full, effective, and sustained implementation of the Conventions," establishes both communication and instrumental roles for the project in relation to its objective. The communication role of the EGI stems from its ability to help connect components of gender and the environment by providing information and data through indicators on governments' performance translating the gender and environment mandates into national policy and planning. The EGI will also play an instrumental role in helping decisionmakers evaluate progress and make changes necessary for meeting the requirements of the mandates.

The development of the EGI exemplifies one of many ways an index can be initiated and how an index can have many roles related to its objective. Additionally, this project demonstrates the need to identify a target audience and users when establishing the foundation of any index framework, because indeed the EGI aims to present information to its users as both a guide and a communication tool for important decisionmaking.

▶ Issue Framing

- Combine complex or disparate data/information to communicate something new.
  - For example, plans by the International Union for the Conservation of Nature (IUCN) to develop an Environment and Gender Index (EGI) that combines information on performance related to both gender and the environment to develop a new metric that is a measure of both (see *Box 2.1*).

▶ Decisionmaking and Management Tools

- Provide a guide for allocating resources, taking actions, or making decisions.
  - Manage environmental problems, create an action plan, or determine a general policy/management direction.
- Devices for support or influence.
  - Build learning and cooperation toward a policy debate or management solution.
- Inform specific decisions or directions of management/policies.
  - Help develop goals, objectives, and tools for policy or management processes.

Our experience in developing both the ESI and the EPI has revealed the importance of a clearly defined overarching goal for the design and communication of an index. As mentioned in Chapter 1, the ESI grew out of an identified need to provide a metric for sustainability analogous to GDP. Key components of the ESI were derived from the long-term aim of developing a single measure to communicate 1) the status of a nation's environmental systems; 2) the pressures on those systems; 3) human vulnerability to environmental change; 4) a country's capacity to respond; and 5) contributions to global environmental stewardship. These five objectives were identified as "components" to more narrowly focus the broad concept of sustainability for which measurable variables could then be identified (see Chapter 4 for more details on defining measurable objectives).

When we transitioned from the ESI to the EPI, our goal changed. Rather than providing a relative measure of sustainability for countries, as in the case of the ESI, we changed the focus of the EPI to measure country performance against absolute targets for which countries could take policy action. We also wanted to highlight issues of environmental data availability, measurement, and quality using the best data available globally but also identifying gaps in understanding. The latter goal has become more prominent in the latest iteration of the EPI, which avoids reliance on modeled estimates of environmental indicators and therefore includes fewer indicators than the 2010 EPI (see Chapter 5). *Table 2.1* summarizes the

Category	2005 ESI	2006 EPI	2008 EPI	2010 EPI	2012 EPI & Pilot Trend EPI
Objective	Gauges the long term environmental trajectory of countries by focusing on "environmental sustainability."	Assesses current environmental conditions.			Assesses current environmental performance and makes comparisons of environmental conditions over time.
Design	Provides a relative measure of past, current, and likely future environmental, socio-economic, and institutional conditions relevant to environmental sustainability.	Provides an absolute measure of performance by assessing countries on a proximity-to-target basis.			Provides measures of performance by assessing countries on a **proximity-to-target** basis over the last decade to track changes in performance over time.

Table 2.1. The evolution of the ESI and EPI's objectives and framework over time.

differences in objectives between the ESI and EPI over the last decade, as well as the influence of these changes on the overall design of the indices.

In delineating the goal for an index, it is helpful to examine an index's purpose in the context of **scope** and **scale**. We define scope as the extent of the area or subject matter considered in an index (i.e., water sustainability versus environmental sustainability). We define scale as an index's range (i.e., project-based, local scale versus national or global). For example, data may be easier to collect for an index that is tightly focused and spans a small geographic area. It can also translate into a more narrow definition and selection of indicators. An example of this type of index could be a company desiring consistent metrics or Key Performance Indicators (KPIs) for its immediate operations within a region. On the other hand, defining scope and scale too narrowly may be limiting and thus unintentionally exclude potentially interested users. For example, an index focused on municipal waste in one particular city may not be broad or inclusive enough to influence state or national decisionmakers who are interested in regional waste concerns. *Box 2.2* describes an example of Armenia's efforts to include a sustainability component in the Human Development Index (HDI) based on their goal of expanding its scope to include environmental concerns.

Box 2.2. Adding a Sustainability Dimension to the Human Development Index

Armenia has been working since 1995 to transform the Human Development Index (HDI) into a Sustainable Human Development Index (SHDI). The HDI attempts to create a summary measure of human development across three basic dimensions: health, education, and income. The HDI uses a single statistic to serve as a frame of reference for a country's social and economic development. It sets a minimum and maximum for each dimension, called "goalposts," and then gauges where each country stands in relation to the goalposts, normalized as a value between 0 and 1.

Using these same principles, Armenia set out to incorporate an environmental sustainability dimension into the HDI. Environmental indicators in the SHDI are divided into two types: those relating to the environmental state of a territory, and those relating to the environmental evaluation of human activities. These two groupings of indicators are then further subdivided into 11 metrics that relate to specific measurable environmental issues.

The structure and indicators included in Armenia's SHDI resemble the indicator framework of the YCELP-CIESIN EPI, although taken together, the SHDI is more similar in breadth to the ESI. The SHDI and EPI represent two indicator frameworks and indices that are designed to serve different objectives and are based on different principles. While the SHDI's aim is broad and encompasses as many aspects of sustainability and human development as possible, the EPI's aim is more narrowly focused. A major principle of the SHDI is to be widely inclusive, while the EPI is based on a sharp focus on performance-related indicators.

Both of these indices demonstrate the possibility of incorporating dissimilar objectives and underlying principles into a similar framework, and how many different elements can be examined and utilized to tailor an index toward a specific scope or goal.

Audience

Another key question to ask in the design of an environmental performance index is, “Who is the index intended to reach and influence?” Defining a target **audience** will help sharpen the scale and scope of an index, as well as facilitate the design of the indicator framework, indicators, and data selection. Setting a clear boundary for an index’s audience can help to act as a filter for what is pertinent in the selection of data and indicators (see Chapter 4 for more details on data selection).

Key questions that can help to define the target audience for an index include:

- Who is the index intended to reach or influence?
- Who will use the index, help disseminate the index, or share the information provided by the index?
- What are the goals of an index’s target audience?
- Should the index be geared toward a general audience or a specific sector of the population?
- Who may be interested in the results of the index, and who is the project relevant to?

An index can be designed to influence a variety of audiences, which can broadly include policymakers, the general public, and businesses. It can also include more specific sectors in these categories: for example, environmental policymakers, or a subsection of the population interested in climate change. In the case of the World Wildlife Fund (WWF)'s Environmental Paper Company Index, the target audience was the pulp and paper industry, which WWF hoped to influence through increasing consumer awareness using the index. Another example is seafood sustainability indices such as Monterey Bay’s Seafood Watch rating or the Global Aquaculture Performance Index[7], which is targeted toward influencing seafood consumers to make sustainable choices. *Table 2.2* describes the objectives and target audiences for several other index efforts.

Index Name	Created By	Goal	Audience
Green City Index	Siemens	To measure the environmental performance of 120 cities around the world.	City and national-level governments.
Multidimensional Poverty Index	United Nations Development Programme (UNDP)	Records which aspects of poverty are felt by different communities around the world in an attempt to make the concept of poverty more easily understandable to policymakers.	Country-level policymakers in government and NGOs seeking to alleviate poverty.
Happy Planet Index (HPI)	New Economics Foundation (NEF)	The goal of the HPI is to develop an alternative measure of human well-being and progress.	Country-level decision-makers and policymakers; the United Nations.
Environmental Paper Company Index	World Wildlife Fund (WWF)	Drive the pulp and paper industry to be more sustainable in its operations; to inform consumers of the ecological footprint of some manufacturers.	Can be used both by consumers in making purchasing choices and by the paper companies themselves.

Table 2.2. Examples of index efforts and their identified objectives and target audiences.

[7] Volpe, J.P., M. Beck, V. Ethier, J. Gee, A. Wilson. 2010. Global Aquaculture Performance Index. University of Victoria, Victoria, British Columbia, Canada.

Principles

What criteria will guide the development of an index? In the planning phases of an index, it is often helpful to identify guiding principles in conjunction with defining an index's objectives and audience. These principles can also act as further filters to aid the developers in gauging what is relevant for inclusion in an index. Clear principles can also help to establish the basis of an index and provide support for the reasoning behind its processes, therefore increasing the transparency of an index for a target audience.

When considering which principles to adopt to guide the construction of an index, one can ask what tenets (e.g., transparency, accuracy, etc.) are important to the developers, users, and audience. Some principles to consider for adoption include:

- **Comparability**

  An index may aim to achieve the greatest basis for comparison as possible, which may narrow the consideration of common indicators and datasets.

- **Accuracy**

  Developers of an index should strive for accuracy—the degree to which the information conveyed in an index correctly estimates or describes the properties it intends to measure—as a principle. An index is only as credible as the underlying data from which it is based.

Box 2.3. When an Index's Goal plays a Critical Role: The 2012 EPI's Water Quantity Indicator

For the 2012 EPI, our team debated whether to include a less than ideal water indicator or to altogether omit a water policy category from the Ecosystem Vitality objective. None of the globally available water data met our established criteria (see *Box 4.1* for a full listing of the criteria used in the 2012 EPI). We were then faced with a decision to exclude the consideration of water resources from the Ecosystem Vitality objective, which would send a strong signal as to the inadequacy of global water data, or to find a sub-par dataset that did not meet the data selection criteria.

The EPI's goals, target audience, and principles guided us to the decision to include a water indicator, even if the indicator did not meet all the criteria we initially set for inclusion. Because water is an integral part of environmental sustainability and natural resources, we decided we could not release the EPI without the inclusion of a relevant indicator to measure water performance, despite no dataset fully meeting our evaluation criteria. The only indicator we found to partially meet some of our criteria was a measure of the Change in Water Quantity, although this indicator lacked a historical time series and relied on modeled data using sparse underlying observations. However, we extensively reviewed alternatives and determined that this dataset represented the "best available," recognizing that the indicator is still only an approximation of overall water resource management regarding ecosystems.

▶ Flexibility

An index may want to adopt a principle of flexibility to allow for possible changes or updates to the index in conjunction with advances in science, data availability, or thinking. Flexibility can also be manifested in the weighting and aggregation process. An index can allow flexibility by providing users an option to customize statistical weightings for a particular purpose. The 2008 EPI provided this option in a web format.

▶ Transparency

It is best practice to adopt a principle of transparency to ensure that data and indicator selection, as well as calculation methodology, are credible for users and key audiences.

▶ Completeness

Developers may also want to adopt a principle of comprehensiveness to help guarantee the index covers all relevant issues. Are there critical aspects of an issue that are missing from the framework? If so, can we justify their exclusion? Completeness may be a critical principle for users or the intended audience of an index, as overlooking an indicator or aspect of an issue may affect an index's credibility. *Box 2.3* describes the important trade-off between completeness and accuracy that we considered for the EPI when debating the inclusion of a water resources policy category.

Conclusion

The main purpose of an environmental performance index is to communicate a goal to a key audience in a credible manner based on a transparent framework of indicators. Toward this aim, this chapter has outlined the importance of developing an index's foundation—the overall goal, target audience, and principles that underpin an index. It is important to keep in mind these foundational elements when moving through subsequent chapters, as they will help index developers make decisions regarding the inclusion of data, indicators, and methodologies.

3 Preparing the Process

Topics covered in this chapter:

- **Management Team**
  Organizing a core team with specified roles to effectively manage the process of developing an index.
- **Engaging Experts**
  Collaborating with experts to share knowledge and receive input and feedback on the index construction process.
- **Process and Timeline**
  Establishing a productive process and approximate timeline for creating an index.

Introduction

The organization and composition of a team developing an index can be highly varied – from a single individual to an inclusive multi-stakeholder process that involves multiple members. In the case of the EPI, we relied on the engagement of multiple institutions, stakeholders and experts. However, it is important to evaluate other factors that contribute to the overall structure and goals of an index, such as the organization and establishment of a management team and the production process. These considerations are heavily dependent upon the type of index being built, its level of complexity, and the type of data relied upon or pursued. Here, we demonstrate the experiences of the Yale and Columbia team in developing the ESI and EPI as a model for understanding these important factors and guiding the establishment of an effective process for the development of an index.

Management Team

The management structure for developing an index is flexible and is largely determined by available human and financial resources. While involving various actors is an essential part of an index development process, it is the **"core team"** that manages the actual process of constructing an index. Because the EPI is a joint initiative between two academic institutions, our management process is somewhat more complex than when there is a single institution spearheading an effort. A more democratic management process with multiple leading institutions could mean greater inclusivity but may also require more time to gain consensus. The model provided in this section may be streamlined depending on the scope of an index and the size and capacity of the creating organization(s). Environmental performance indices typically encompass a broad range of issues, and it is likely that a management team will need to draw from outside entities to be properly inclusive and thorough when developing the index.

Generally speaking, the core EPI team has consisted of one or two project "vision leaders," a project manager, research staff, analytical staff and administrative support. A vision leader provides the overall direction for an index and is usually an expert in the targeted field of the index. To effectively guide the development of an index, the vision leader must be knowledgeable of the context for which an index is being constructed. The EPI relies on the expertise and experience of vision leaders who are well-versed in the language of policy and its drivers, as well as related environmental data and trends. The vision leader also plays a critical role in resolving disputes during the development an index. For example, if there are issues surrounding the inclusion of certain policy categories or indicators, the vision leader can make a final decision. For the EPI, the vision leaders are not necessarily involved in the day-to-day research tasks but instead serve a more advisory role in providing consultation and overall guidance. The role of a vision leader could easily take the form of an advisory or consultancy board of people, or instead be driven by stakeholder input.

Other members of the core team are often more involved in the day-to-day management of an index. The project manager serves as the single point person organizing the details of the project. The project manager helps define goals, assign and track tasks, establish timelines and provide feedback on work at all stages of the project. In addition, the project manager provides input for the content of the index and the preparation of the report.

The research staff conducts preliminary research, identifies key **experts** for engagement (discussed later in this chapter), provides guidance on indicator selection and development, and assists in data collection, processing (including formatting as well as transformation of spatial datasets using Geographic Information Systems (GIS), etc.) and interpretation. For the analytical and statistical aspects of constructing an index, core team members with quantitative skills or statistical capabilities provide the bulk of data exploration and analysis, as well as indicator and index construction (see Chapter 6). Organizations that

Figure 3.1. Summary of key roles for core team members who construct the EPI. The core team structure suggested here can be tailored and modified to fit the scope and scale of an index. Often external consultants or university partners may assume the roles of a core team or specific members if an institution lacks the internal capacity to fulfill each role.

lack these necessary skills may choose to hire an outside consultant to assist in the process of developing an index. However, it is valuable to have at least one team member with the quantitative skills necessary to assess or cross-check the validity of calculations and results. It is also essential to have administrative staff to support each phase of the project. Administrative support members help manage resource and staff requirements, organize meetings and workshops, and take on vital tasks such as website administration, vendor liaising, and budget management. *Figure 3.1* summarizes the key roles of each core team member.

The Contribution of Experts

In addition to a core team, a wider network of experts can be convened at several key stages to provide feedback and guidance on methodology, indicator and data selection, and preliminary results.

The right experts for an index help facilitate stakeholder buy-in and adoption.
Experts are often "thought leaders" whose knowledge is valued by the stakeholders of an index. Their

input helps ensure an index provides the most credible information and builds confidence among stakeholders to promote widespread adoption and acceptance of the results. With this in mind, three types of experts can add to the robustness of an index:

1. **Subject area experts** check for the validity of data and determine what the data is used to measure. They understand the fundamental environmental issues that a given indicator or category is attempting to measure, have domain expertise (including knowledge of existing datasets and monitoring systems), and can directly contribute to or assist in the selection of appropriate data for the purpose of indicator construction.

2. **Methodological experts** contribute the latest knowledge and methods of index construction. They can often provide valuable options and help guide the appropriate selection of methods.

3. **Target audience experts** provide a preliminary reaction from a target audience to ensure the project is producing something relevant for its purposes. They are also knowledgeable on the needs of specific target audiences and can offer valuable suggestions for meeting those needs.

Experts help shape the overall design of an index.

From advising on equivalents of "policy categories" to helping vet datasets and selecting indicators that provide the best signal for what an index seeks to measure, experts can guide efforts for the best possible result. In the realm of data selection, experts are often aware of what data are commonly used and trusted, as well as data sources that may not be regarded as reliable but are in the public eye. When the time for data analysis comes, experts can advise on almost everything – from defining the meaning of zeros and negative numbers in particular datasets, to determining whether anomalies are true or mere coding errors. Later in the index development process, after the core team completes initial index calculation, experts can serve as "beta testers" by reviewing the index before the results are released to the general public. Because anomalies and flaws in index construction frequently become evident only after an initial analysis has been completed, experts play a crucial role in the beta testing stage by exposing shortcomings and troubleshooting issues before others begin to use the index. From start to finish, experts bring a depth of learning to a team and complement the breadth of knowledge.

Experts can help identify the critical issues, key data and indicators within a field.

Our experience with the EPI has shown that experts are aware of the nuances of their field in a way generalists are not. At an early stage in the development of the 2012 EPI, our core team circulated a list of selection criteria, along with a list of potential indicators, to a range of subject area experts. We asked these experts to rate the proposed indicators against each criterion and identify new potential indicators that met those criteria. This feedback gave our team valuable leads early in the process. For example, when looking for fisheries data, experts from the *Sea Around Us* provided our team the most relevant, up-to-date datasets by which to construct the best indicators for fisheries sustainability. These experts provided guidance on one dataset, the Aquaculture Performance Index, which we found difficult to aggregate to a country-level metric and was excluded from further consideration. They also helped us evaluate the Marine Trophic Index and noted the need to interpret the data in conjunction with the extension of fisheries off of the coastal shelf; essentially making the data impractical for our purposes. The *Sea Around Us* Project experts contributed indicators for the Percentage of Stocks Overexploited and Collapsed by Exclusive Economic Zone (EEZ) and Trawling Catch per EEZ, and this collaboration enabled us to improve the fisheries indicators in the EPI so policymakers had a more accurate, relevant picture of the status of their marine ecosystems.

Experts can contribute their own data and indicators to an index.

Several experts directly contributed data for the EPI. The *Sea Around Us* Project processed their fisheries data, along with the time series, to provide country-by-country data consistent with the EPI methodology.

This saved our team significant time and potential errors, because often raw data must be transformed (e.g., statistical transformations or unit conversions) to better provide transparent and applicable information. In other cases, experts may provide raw data that can be transformed by members of the core team. For example, an expert at the University of Maryland provided satellite-derived deforestation maps, which were then used by CIESIN to calculate the percentage of forest cover lost in a GIS. We then used this data for the 2012 EPI's Forest Loss indicator. The Battelle Memorial Institute derived countries' satellite-based PM2.5 data for the Air Quality – Human Health effects policy category by normalizing the satellite-derived fine particulate matter concentrations with population data to determine a country's average exposure. By convening these experts for input on the EPI, we were able to streamline the index construction process and produce a much higher quality project with fewer resources.

Experts can help cross-check and validate data.

It may be difficult to verify data with a relatively small core team, particularly when the scope of an index is global like the EPI. Therefore, for the EPI we rely on subject-area experts who are more familiar with the data used to help explain any anomalies or inconsistencies in the data. Experts at the *Sea Around Us* Project informed us of underlying problems within several fisheries datasets that we may not have otherwise understood. For example, the data for Greece included a suspicious spike in improved performance in the year 2004 in the time series Trawling Catch per EEZ. Through consultation with the *Sea Around Us* Project experts, we discovered that there was a known reporting error for this indicator in 2004. This led us to exclude the data point from the time series and **"smooth"** Greece's data by creating an average value based on the previous and subsequent years' data. Working with and building a network of experts gave us the ability to cross-check and verify individual datasets, lending greater accuracy that would be otherwise difficult to achieve.

> Working with and building a network of experts gave us the ability to cross-check and verify individual datasets, lending greater accuracy that would otherwise be difficult to achieve.

Experts can help provide added context and detail for a dataset.

For the EPI's renewable electricity generation indicator, we wanted to ensure we were not rewarding countries that burned crop waste or animal dung. Although these resources are technically renewable, burning crop waste or animal dung produces air-pollution by-products. We consulted with experts at the International Energy Agency (IEA) who informed us that data for renewable energy consumption usually include these biomass measurements. Upon request, the IEA experts were able to provide us two separate datasets for renewable electricity generation: one including a biomass portion and one without. Although the difference between the two datasets turned out to be negligible, the transformation of the data into two distinguishable datasets prevented any potential objections to the indicator by EPI users. As exemplified here, expert contextual input can help point the team toward the most salient data for reaching specific goals.

Evaluating Expert Input

Often, expert involvement and input can have the effect of adding noise to the development of an index, particularly when expert opinions present conflicting results. It is the core team's role to evaluate expert judgments and decide when to accept an expert's opinion. There are also more objective solutions for deciding how to weigh expert opinion in the construction of an index. Chapter 5 describes how to gauge expert inputs for the quality of datasets as an example of a more quantitative assessment (see *Appendix 4* for the process of grading data quality for the 2005 ESI).

Regardless of how expert opinion is used, it is critical to properly attribute and credit experts' contributions to the index, both for ethical reasons and so experts will feel a sense of ownership and be willing to

contribute to future editions. We have worked with experts to co-author scientific publications based on their collaboration. The Joint Research Centre of the European Commission has also partnered with the EPI to conduct independent sensitivity analyses to understand uncertainties in the EPI methodology. In some cases, we have also provided honorariums to compensate experts for their participation.

Process and Timeline

The construction of past EPIs generally required a year to complete. However, the timeframe for constructing an index can vary depending on the starting point and existing foundations in place prior to processes for data collection, indicator calculation, and review. The first attempt at creating an index may require more time. Chapter 2 presents in detail the process of laying the foundations for an index, as described in step one in the process outlined below and summarized in *Figure 3.2*. It is also critical to note that experts can be engaged in every step of the process. However, it may be advantageous to the core team to structure feedback processes efficiently by aligning expert reviews at key stages.

A generalized timeline for the EPI can be expressed as follows (detailed explanations for each step are described in other chapters, as noted in each step):

1. Define Overall Goals, Audience, and Principles (Chapter 2)

The first step in creating an index is laying its foundations. At this stage, preliminary research and review of scholarly literature is helpful in establishing a scientific basis for an index. Here, the 'why,' 'for whom,' and 'how' are defined.

This phase of the index development process can require a substantial amount of time, depending on the core team's timeframe to clearly define an index's goals, audience, and principles. One strategy to make this

Figure 3.2. A generalized timeline for the construction of the EPI.

process more efficient is to narrowly focus an index on a specific issue, target a certain audience, and clearly elaborate principles by which to guide the process. A narrower scope will also help to ensure the rest of the index process remains on track.

2. Seek Preliminary Expert Engagement (Chapter 3)

It may be helpful to engage experts and stakeholders to provide guidance and feedback for Step 1 and, for later steps, help a team understand how the index fits into the larger landscape of stakeholder concerns and research in the area. At this stage, the team should define the scope of expert engagement and identify experts based on areas of expertise.

Based on the overall goals, audience, and principles of the index, a list of the types of experts needed for the project can be defined. Types of experts who can be engaged in this stage of the index development process are described in the above section, The Contribution of Experts.

Once the expert contributors have been identified, they can begin reviewing the overall goals, audience, and principles established in Step 1. Their feedback should include advice on the current project scope and details, as well as how to reach the project's goals. They can also suggest datasets potentially relevant to the project and the relative importance of issues encompassed by the index (e.g. air pollution vis-à-vis forest cover). Furthermore, they can help to identify other experts to engage in the process.

3. Establish Indicator Framework (Chapter 4)

After defining an index's objectives, goals, and audience, the core team should work to define the indicator framework that will determine the categories the index will measure, as well as possible indicators to gauge performance in those categories.

As detailed in Chapter 4, establishing the structure of an index is an iterative process that may require several rounds of revisions. Often an indicator framework starts from an "ideal" perspective, including the categories and indicators that a model framework would include discounting any data limitations, but is then refined with stakeholder and expert input. Establishing the EPI indicator framework, which has been a relatively fixed process since 2006, has proven to be one of the more time-intensive stages; requiring slightly less time than the collection of indicator data. Often, we required many weeks and multiple rounds of feedback to configure and finalize the final framework.

4. Collect Data (Chapter 5)

This step can be the most time consuming part of the index construction process, depending on the sources for data collection.

Researching and gathering data for the purposes of index construction is an important step that should be conducted thoroughly and systematically. However, specific details as to how to go about collecting appropriate data for the construction of an index is outside the scope of this manual (users can refer to *Appendix 1* for these resources). As further explained in Chapter 5, the underlying methodology and sources for data should be carefully evaluated. The evaluation of consistency and trends in underlying methodology and data sources will often elucidate issues of data quality. It is important to consult with experts when collecting and analyzing data, and their comments and suggestions are helpful throughout this process. Many datasets will fail to meet specific criteria set by an index team and may be excluded from the project. However, it is best to retain any data collected because it may be helpful in assessing other data sets, detailing missing information between sources, and interpreting improvements in future iterations of the published data. It is also useful to communicate with the sources for data collection. Often, the publishers of data are looking for feedback to improve their products or may suggest more suitable sources or datasets for the purposes of an index.

5. (Optional) Conduct Expert Review on Framework and Indicators (Chapter 3)

At this stage in the indicator construction process, the core team can choose to engage expert feedback on an index's framework and indicators. Experts can build on their preliminary thoughts on the foundations of the index to give a more in-depth evaluation of the framework and datasets with regards to established criteria. They can provide guidance on the inclusion or exclusion of data sources, the weighting of indicators, and indicator construction.

To prepare for this round of feedback, the core team provides experts with a description of the index framework, information on selected datasets (including the background on each and the reason it was chosen), and a means of communicating feedback, whether electronically or verbally. The experts should be given ample time (i.e., several weeks) to carefully complete this review. Although experts may be encouraged to focus only on their area of expertise, their comments on other areas may also be beneficial. Because many environmental subject areas overlap, experts can provide valuable input and suggestions for inter-related topics.

Once this round of expert review has been completed, the core team aggregates feedback, solidifies the index framework and weightings, and selects the final datasets to be used.

6. Incorporate Feedback and Calculate the Index (Chapters 3, 5 and 6)

At this stage, the core research team processes the data through the chosen methodology to produce the index. Anomalies and any questions for the final round of expert review are noted.

Research staff members prepare the selected datasets following the expert review, ensuring that all datasets are in a compatible format for statistical analysis. This includes consistency in notation for missing values and units of measure (see Chapter 5). Next, the cleaned datasets are used to produce the overall index results (see Chapter 6).

Once the preliminary results have been prepared, the team should create a detailed presentation of these results for the next round of expert feedback. This presentation should note anomalies, missing data, and places where the team is seeking help in balancing trade-offs and making a final decision.

7. Hold Expert Review for Feedback on Results (Chapter 3)

For this expert review, the core team presents the results of preliminary analysis and invites questions and comments on the outcome. This review reveals any conspicuous oversights and provides feedback on the overall results and outcome of the index.

A final round of expert review should be held when the draft of an index is nearing completion. For this round of feedback, we always find it helpful to convene experts for a review workshop. Experts are given a draft of preliminary results prior to the workshop so they can comment and formulate initial feedback. It may be beneficial to divide experts into subject area teams before the workshop so they can begin discussing key issues in preparation for meeting with their group.

At the workshop, experts participate in a presentation of the results and a roundtable discussion on the overall outcome. Next, subject area teams can meet to provide more in-depth feedback on results. The results of these breakout groups are communicated to the core team and entire expert group to share lessons across subject areas and reach a consensus on tough questions.

8. Report and Outreach Preparation

At this stage, the core team prepares the index report and any additional communication tools, such as a website, as well as a media and target audience outreach strategy.

The EPI report goes through several rounds of internal review and, once finalized, includes detailed objectives, methods, indicator framework, data sources, and results of the index. A comprehensive outreach strategy, including a website, should be considered far in advance of drafting the report and should help to maximize the reach of an index.

9. Report Release

For the purposes of keeping a project on track, it may be helpful to time the release of an index with a conference or other event.

We have traditionally partnered with the World Economic Forum (WEF) to release the biennial EPI report at its Annual Meeting. Timing the launch of the report with the WEF meeting has provided a hard deadline for the release of the EPI. Additionally, this release strategy and venue provide an international forum by which to disseminate the results.

10. Measuring Impacts

A last key step to developing an index is to measure its impacts. It is important to determine if the index has reached the desired audiences and has exhibited the intended impacts.

Taking note of media mentions, website hits, and report downloads, as well as scheduling one-on-one interviews or focus group meetings with representatives of the target audience, can help to measure the impact of an index. This important last step should not be overlooked when developing an index, because it can also help to provide valuable feedback for incorporation in future editions of an index.

Conclusion

The process and stakeholders involved in developing an index are just as important as indicator selection, calculation methodology, and statistical aggregation (discussed in Chapters 4-6). However, it is important to keep in mind that the process and steps outlined in this chapter are by no means linear and are in fact more iterative. Therefore, while we intend for the discussion or explanation of these modes and processes to serve as recommendations, we recognize that our approach is not a "one-size-fits-all" method that will work for everyone. We simply have refined our collaborations over time and feel that our experience may be of value to others when planning similar efforts. The timeframe for creating an index is highly variable and depends on the experience and organization of the managing team. The template provided in this chapter is only one example of the order and timing we have used to develop the ESI and EPI.

4 Building an Indicator Framework

Topics Covered in this Chapter:

- **Designing the Indicator Framework**

  Measurable Objectives: Establishing an index's goals that can be defined in measurable terms.

  Core Categories: Refining objectives to help group related indicators.

  Indicators: Defining and selecting indicators.

- **Considerations for Indicator Selection**

Introduction

The **indicator** framework is the backbone of an index. It includes the major elements that define an index, such as the measurable objectives, policy categories, and indicators, as utilized in the EPI. A well-established indicator framework aids in the interpretation of a set of indicators and can provide a conceptual map for determining relationships between different issues. Additionally, the indicator framework is essential for making the connection between certain observed trends in data and the actions or policies that may have led to these developments.

While there are many different approaches to constructing an indicator framework (*Appendix 2*), this chapter will only focus on the EPI's methods, which employ three tiers of weighting and aggregation to assign percentages for the purposes of statistical combination and produce a final, composite number.

Designing the Indicator Framework

There is no single approach to constructing an indicator framework. The OECD has published extensive and diverse literature on designing frameworks based on its vast experience in developing measures of environmental performance and assessment (see *Appendix 1*). Many frameworks are built upon "cause and effect" logic that helps developers of environmental indices identify appropriate components and indicators to measure an index's goal. A common causal framework is the Pressure-State-Response (PSR) model developed by the OECD in 1994.[8] The PSR framework (*Figure 4.1*) is composed of:

- **A pressure variable** that describes human activities or aspects that exert pressures on the environment. This variable represents the underlying cause or causes of a problem, whether this is an existing cause or one created from a new activity or investment. Examples of potential pressures include income growth, trade patterns and activities, energy use, and population growth.
- **A state variable** that describes some measurable characteristic of the environment that results from the pressure variable. Examples of state variables, or indicators, in the EPI include water quality or quantity, deforestation, and air pollution.
- **A response variable** that measures policy responses to environmental states and changes, such as actions or investments that are instituted to address a problem. Responses to environmental problems can affect a state variable either directly or indirectly. Examples include water-pricing methods, the use of alternative crops, and reforestation programs.

One of the criticisms of the PSR framework is that it oversimplifies complex relationships because there are often numerous pressure, state and response variables. Adjustments have been made to the PSR model to create the Driving Force-State-Response (DSR) framework.[9] The pressure variable is replaced with the driving force variable, or the factor influencing a specific activity. The PSR model has also been amended to include impact indicators, making it a Pressure-State-Impact-Response (PSIR) framework. Adding both Driving Forces and Pressure variables transforms the framework into Driving Force-Pressure-State-Impact-Response indicators (DPSIR).[10,11] *Figure 4.2* illustrates the relationships between these various frameworks in detail.[12]

---

8 Organisation for Economic Development and Cooperation (OECD). 2003. OECD Environmental Indicators: Development, Measurement, and Use. OECD. http://www.oecd.org/environment/environmentalindicatorsmodellingandoutlooks/24993546.pdf

9 United Nations Commission on Sustainable Development (UNCSD), 2001. Indicators of Sustainable Development: Framework and Methodologies. Background Paper Number 3. Available: http://www.un.org/esa/sustdev/csd/csd9_indi_bp3.pdf.

10 European Environmental Agency. 2003. The DPSIR framework used by the EEA. http://ia2dec.ew.eea.europa.eu/knowledge_base/Frameworks/doc101182/

11 Convention on Biological Diversity. 2003. Report of the expert meeting on indicators of biological diversity including indicators for rapid assessment of inland water ecosystems. Convention on Biological Diversity, 10–14 November 2003, Montreal.

12 Segnestam, L. 2002. Indicators of Environment and Sustainable Development: Theories and Practical Experience. Environmental Economics Series, Paper No. 89. World Bank. http://siteresources.worldbank.org/INTEEI/936217-1115801208804/20486265/IndicatorsofEnvironmentandSustainableDevelopment2003.pdf.

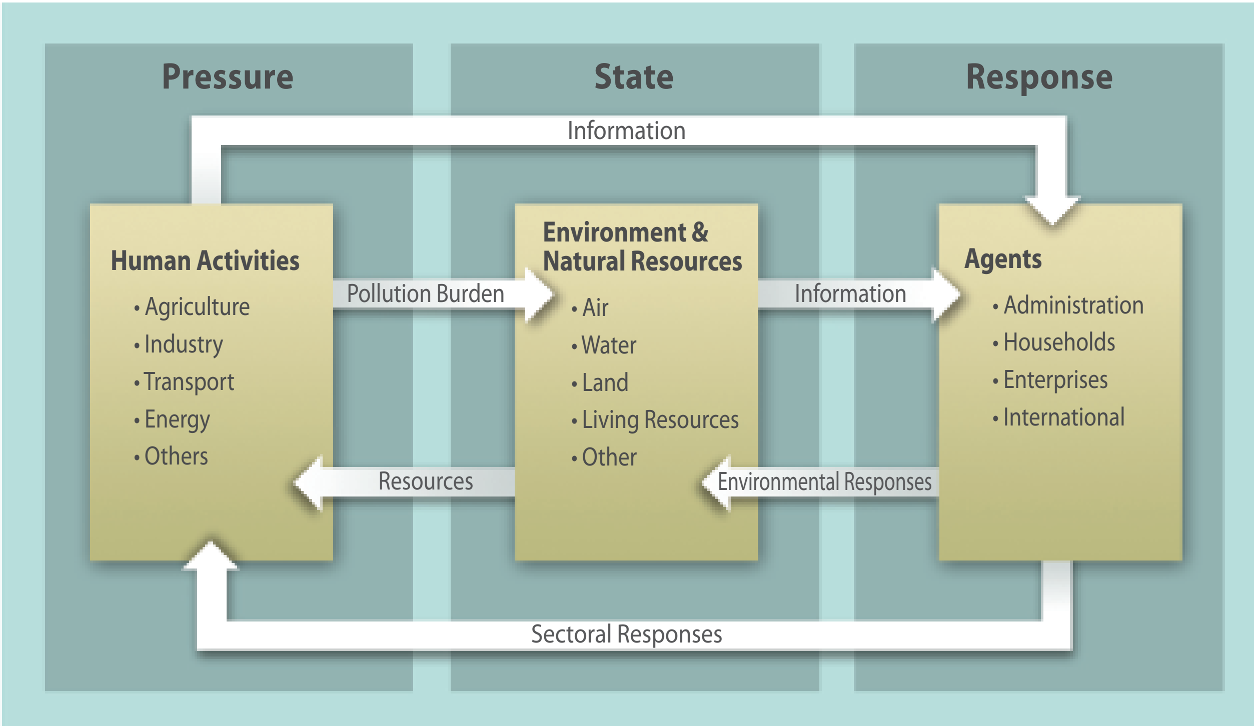


Figure 4.1. The OECD Pressure-State-Response model isolates various elements of the causal chain in environmental issues (adapted from Segnestam, 2002).

It may also be useful to refer back to the scope and the scale of an index, defined in Chapter 2, to design a suitable indicator framework. As described by the World Bank, different analytical levels can require separate frameworks for analysis, including commonly used frameworks such as:[12]

- **A project-based framework** (also described as an Input-Output-Outcome-Impact framework) is often used to monitor the effectiveness of projects to improve the state of the environment. ***Input*** indicators monitor the project-specific resources provided; ***Output*** indicators measure goods and services the project provides; ***Outcome*** indicators measure the immediate, or short-term, results of project implementation; and ***Impact*** indicators monitor the longer-term or more pervasive results of the project.
- **The OECD's PSR framework** and subsequent variations as described above in *Figure 4.2*. These frameworks may be the most appropriate for national, regional, and international-level analyses.
- **A theme-based framework,** such as the use of indicators by the United Nations Commission on Sustainable Development's (UNCSD) for monitoring sustainable development starting in 1995.

We based the ESI and EPI on the OECD's PSR and DPSIR frameworks, which are most appropriate to compare environmental performance at the national and international levels. This is not to say that a similar indicator framework cannot be adapted for smaller scales. There are three major components to the EPI framework: 1) Measurable objectives; 2) Core Policy categories; and 3) Indicators. The following paragraphs will describe a process of developing each component of an indicator framework based on our experience with the EPI.

12 Segnestam, L. 2002.

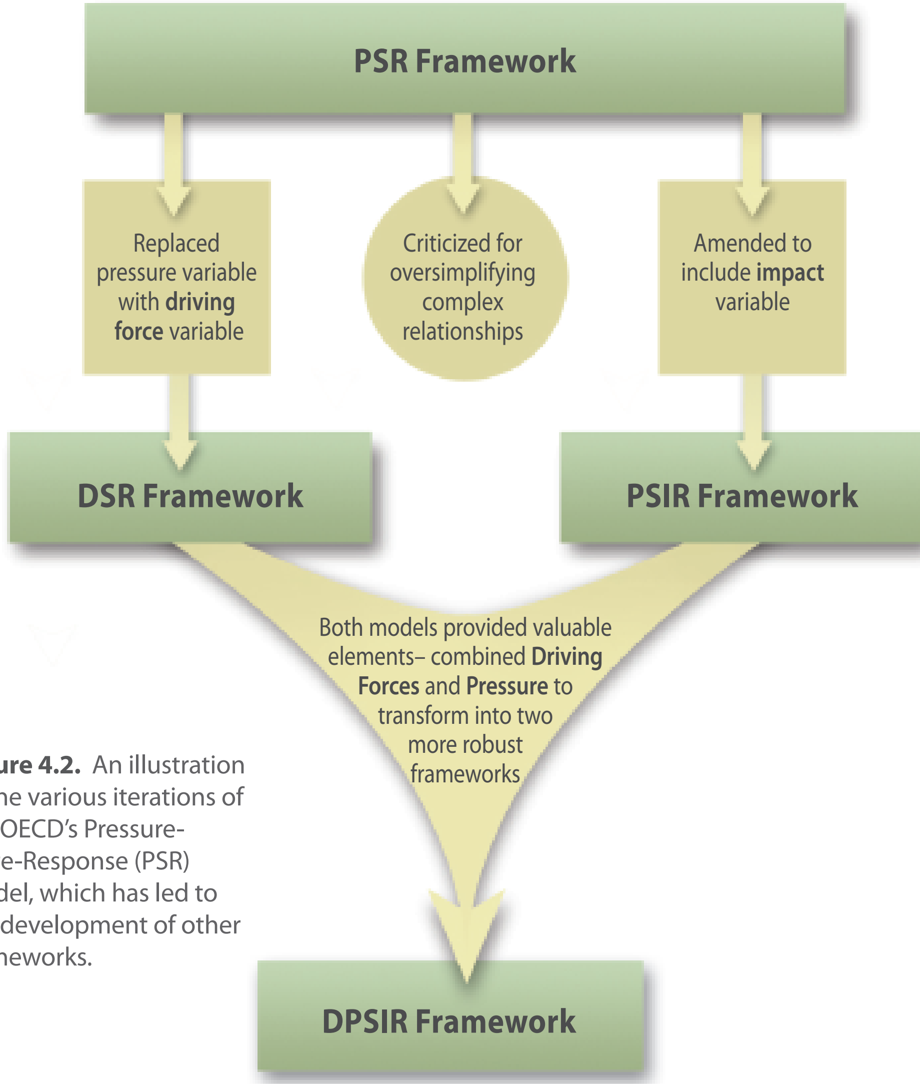


Figure 4.2. An illustration of the various iterations of the OECD's Pressure-State-Response (PSR) model, which has led to the development of other frameworks.

1. Measurable Objectives

As discussed in Chapter 2, a well-conceived foundation of goals, target audiences, and guiding principles underpin a successful index. From this foundation, developers can begin to construct an indicator framework that measures the overall goals of the index. The first major component of the EPI indicator framework is the measurable objective. **Measurable objectives** relate to the overall goals of an index and can be assessed with data. They comprise the top-level of aggregation in the EPI indicator framework. Often, the goal identified for the index (Chapter 2) will serve as the measurable objective, depending on how narrowly it is defined. In the case of the ESI, for example, our goal was to develop a single metric to gauge environmental sustainability. However, sustainability itself was a broad concept that required definition. Therefore, we chose five narrower components, or measurable objectives, by which to delineate our definition of sustainability.

Often it may be sufficient to have only one overall measurable objective for an index. *Box 1.1* (page 7) provides several examples of these types of indices. For example, the Ocean Health Index (OHI) has one overarching goal—to provide an assessment of ocean health. This broad overarching goal is extended into 10 narrower components, or measurable objectives, for assessing ocean health. For other indices, it may be preferable to have more than one objective if the corresponding data and variables used are disparate. The Human Development Index (HDI) measures human well-being in three objectives: health, education, and income. In this case, the identification of three separate objectives allows for clear distinction of objectives within the HDI to make its contents more understandable.

For the EPI, we selected two environmental protection objectives: **(1) Environmental Health -** reducing environmental stresses on human health, and **(2) Ecosystem Vitality -** promoting ecosystem vitality and sound natural resource management. Following a careful review of existing policy goals and literature, we chose these two broad policy objectives because they mirror the priorities expressed by policymakers with respect to the environment and natural resource protection—most notably the environmental dimension of the United Nations Millennium Development Goals, as discussed in Chapter 1. Additionally, the scope of issues represented by Environmental Health and Ecosystem Vitality are defined in measurable ways in terms of existing data, established policy goals and targets, and relevance. *Box 4.2* describes an example of how the EPI framework can be adapted to incorporate additional objectives. In this case of our collaboration with the Chinese Academy for Environmental Planning (CAEP),the EPI framework was modified to incorporate a third objective in line with the policy priorities of the Chinese government.

2. Core Categories

Core categories of an index are the types of categories that reflect facets of its measurable objectives. These categories set clear policy lines to help distinguish an intermediary level between indicators and objectives. Core categories help to summarize the types of metrics used to support measurable objectives and meet the overall goals of the index. While it may not be necessary to identify core categories and add this level of additional weighting and aggregation between the objectives and indicators, core categories may help to organize an indicator framework, particularly when an index includes many types of different indicators. For example, the Climate Change and Energy policy category of the 2012 EPI includes several indicators to capture various aspects of environmental performance on climate change (*Figure 4.3*).

Core categories also help to improve the overall comprehensiveness of an index and indicator framework. For example, policymakers or the public may not understand that carbon dioxide (CO_2) emissions are a measure of environmental performance. Through the introduction of an intermediary definition of "climate change and energy" as a policy category, greater clarity is provided, and a target audience may be more apt to see the connection between climate change and energy emissions and environmental sustainability.

In conceptualizing an indicator framework, the identification of core policy categories may also assist in defining and selecting indicators. After developers take the first step of identifying measurable objectives, core categories can add further clarity to an index by distinguishing elements or components of an objective. For the ESI, the notion of environmental sustainability was inherently too broad to begin at the indicator level. Therefore, we identified five core components—environmental systems, reducing environmental stresses, reducing human vulnerability, social and institutional capacity, and global stewardship—as a means of outlining a clearer definition of sustainability. Each of these categories was chosen because they extended from a broad base of theory in the ecological and environmental policy sciences. After identifying these five components, variables for indicators were then selected. The selection of diverse indicators should relate to a component, but designers of an index may not want to select highly correlated indicators that have

[13] The OECD's (2008) Handbook on Constructing Composite Indicators (http://www.oecd.org/std/clits/42495745.pdf) provides a useful discussion of this issue.

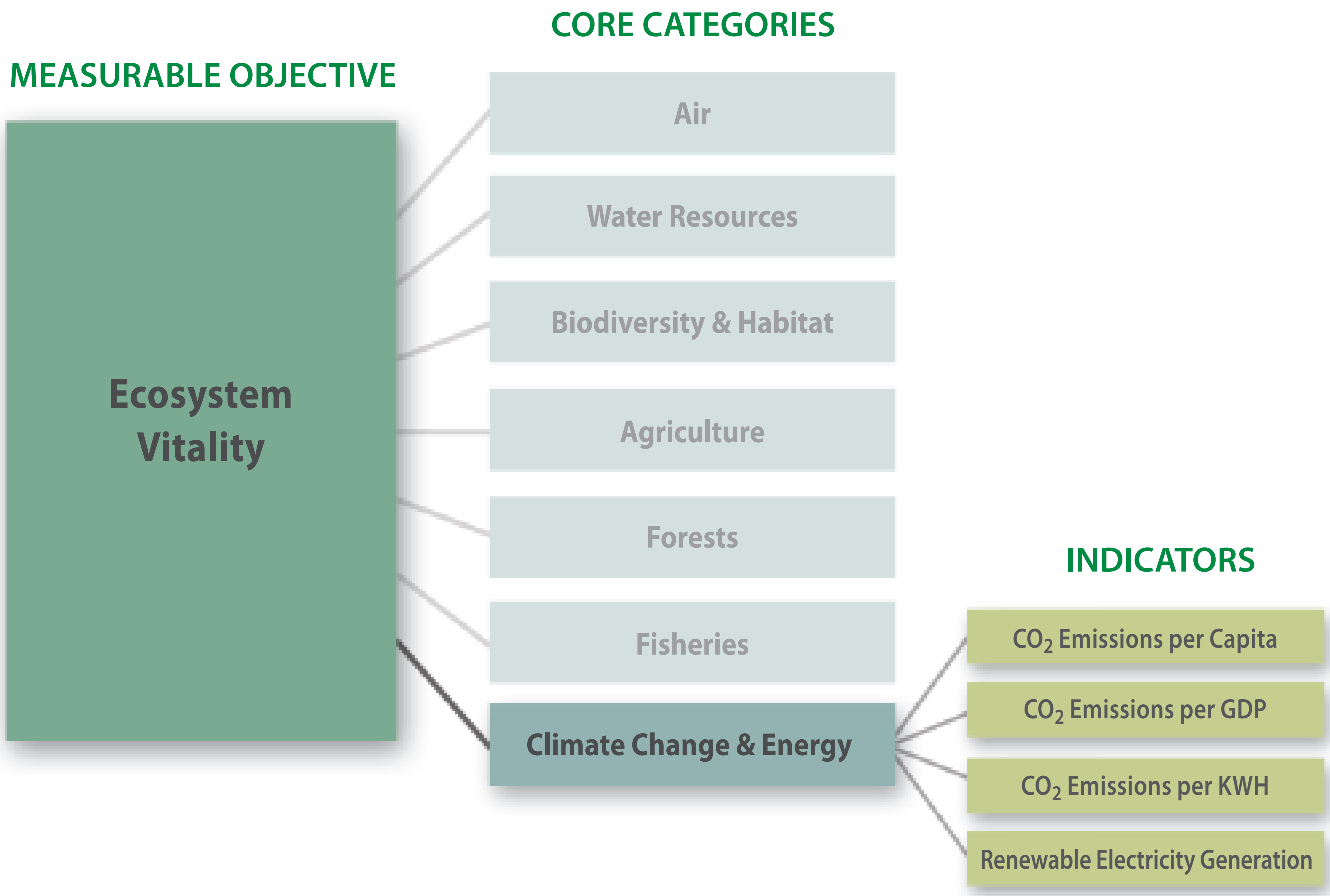


Figure 4.3. The Climate Change and Energy policy category provides an intermediate grouping of climate- and energy-related indicators, and adds an additional level of aggregation.

the effect of adding redundancy to the index.[13] Diverse indicators can also help to provide a cohesive measure of the concept in question. The components not only aided in the categorization and summarization of the 76 indicators used for the ESI, but alone are useful analytical tools to give decisionmakers a sense of how they perform on an environmental issue, such as air quality. Therefore, the components proved to be useful even in isolation from the aggregate index.

The 2006 and 2008 EPIs include six core categories: environmental health, air quality, water resources, biodiversity and habitat, productive natural resources, and sustainable energy. The 2010 and 2012 EPIs include 10 core categories: environmental health, air quality (human health effects), water quality (human health effects), agriculture, biodiversity and habitat, forestry, air quality (ecosystem effects), water quantity, climate change and energy. Overall we chose these categories because they reflect facets of our measurable objectives and are core issues related to environmental performance. For example, we chose Water Resources as one of the six core categories for the 2006 Pilot EPI because water is a critical element for environmental sustainability and is a major topic in the international dialogue on sustainable development.

3. Indicators

The third component of the indicator framework includes the indicators themselves—quantitative variables measured from observations (i.e., data) or calculated. There is no single way to determine appropriate indicators for an index, and the indicators included are inherently subjective for gauging a particular environmental issue or goal. This being said, indicators are not designed to provide the full picture of a particular environmental issue. Instead, they are meant to help reveal trends and draw attention to particular issues (see *Figure 4.4*). Indicators can also characterize various states or conditions, track or predict changes, identify stresses, determine risks, and influence management decisions.

In deciding which indicators to select, the 5D's can help developers determine what types of indicators will help an index convey its message.[14] Indicators can be used to:

▶ Describe issues

- By reducing complexity in policy-relevant ways.
- By answering the question, What's happening?

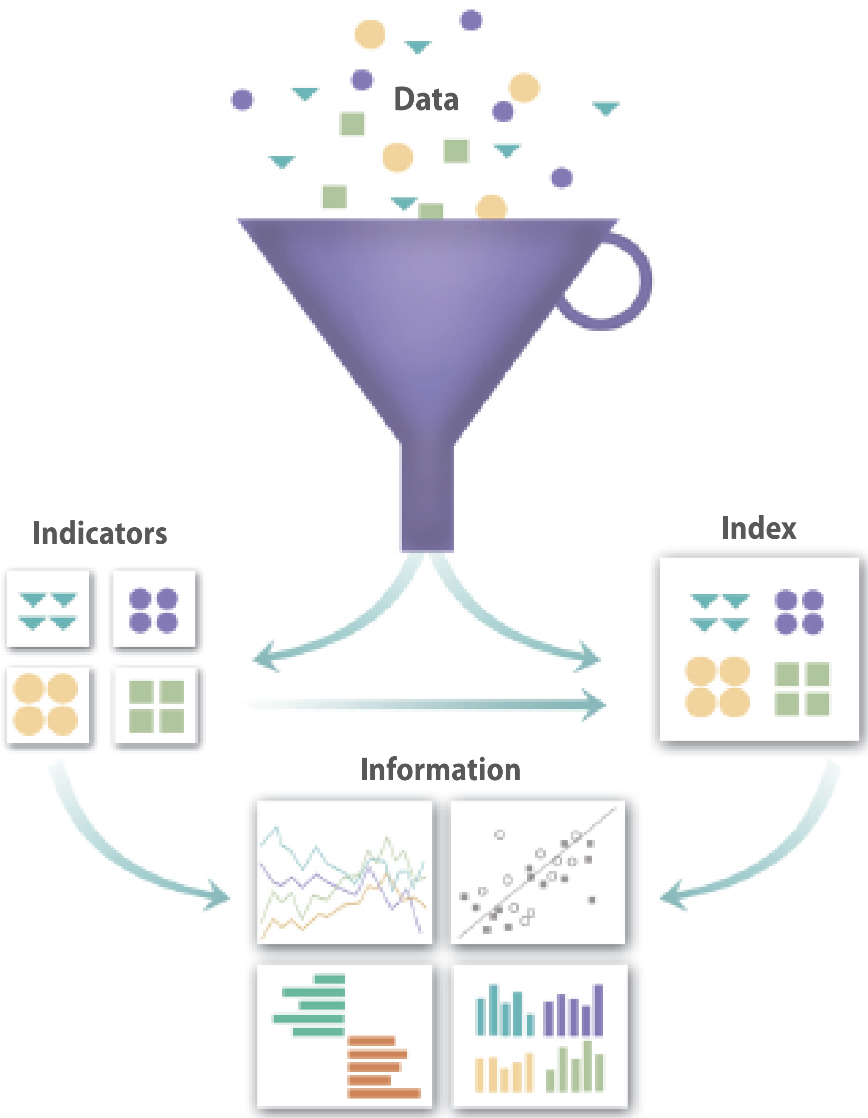


Figure 4.4. A schematic that clarifies the relationship between data, indicators, index, and information. Indicators and indices are tools by which to communicate data in more targeted and effective ways.

14 de Sherbinin, A., A. Reuben, M. Levy, and L. Johnson. 2013. Indicators in Practice: How Environmental Indicators are Being Used in Policy and Management Contexts. New Haven and New York: Yale and Columbia Universities.

- **Diagnose problems**
  - By depicting trends and enabling comparisons between diverse phenomena.
- **Discover patterns**
  - By identifying leaders and laggards.
  - By helping to identify best and worst practices.
  - By targeting resources at stake.
- **Deliberate about solutions**
  - By helping societies and decisionmakers engage in dialogue about what kind of future they want to have.
  - By helping ground discussion in empirical reality.
  - By setting goal posts whose desired positions can be debated.
- **Drive action**
  - By helping to navigate to a desired future.
  - By holding decision-makers and managers accountable.
  - By rewarding progress and punishing inaction.

The types of indicators we used for the EPI framework are performance-based indicators linked to quantitative objectives, such as targets or commitments (described in Chapter 6). Examples of these types of performance indicators include emissions per unit GDP to measure the environmental efficiency of the economy, *per se,* or change in water quantity to measure the intensity of the use of water resources.

Considerations for Indicator Selection

Over the last decade, we've discovered some helpful considerations to keep in mind when developing an index:

▶ Less may be more.

Our experience with the EPI has demonstrated that being selective in terms of which indicators to include often facilitates the communication and understandability of an index. The inclusion of too many indicators may dilute the signal provided by an overall category or goal of an index. It may also be difficult for the target audience to understand the message an index is attempting to convey.

▶ Establish causal connections between indicators and index goals.

In selecting appropriate indicators for a performance index, it is important that the user and target audience can see sufficient evidence for a causal link between the indicator and the issue it is trying to measure. For example, a measure of the rate of deforestation can provide an indication of forest protection efforts, but if one is trying to determine an underlying cause of deforestation, information about the rate alone is insufficient.

▶ If a perfect indicator doesn't exist, look for a proxy.

Due to data limitations, it may not be possible to find an exact indicator to measure a particular issue. When direct measures are unavailable, **proxies** can often be considered as the next available substitute, which may be the case if, for example, a lack of resources prevents direct measurements from being obtained. However, proxies should be carefully considered because they may potentially be misleading. For example, for the EPI we included several measures of protected area coverage—for biomes, highly threatened species and marine areas. These indicators may become confusing to some users if they do not refer to our methodologies and explanation of these metrics. One advantage of an indicator measuring the amount of space or territory under protection in a country is its ability to reveal the types of action being made on the ground, or policy response (i.e., the amount of space or territory covered by protected areas as deemed by a policy decision). But a disadvantage of this indicator lies in its inability to provide information on the substance of an outcome. We cannot know from this indicator, for example, how well a protected area is being managed or whether important species are being adequately protected.

▶ Consider relevant spatial and temporal scales.

Given the lag time for certain environmental phenomena (i.e. reforestation efforts, which may take several years to manifest measurable change), it is crucial to select indicators in an appropriate time period to capture the change ostensibly being measured. For this reason, the trend analyses and change indicators may be more useful in revealing performance rather than simple snapshots of data lacking temporal context. Spatial scale is also an important consideration, one that has proven to be controversial in the EPI. While the EPI provides performance scores at the national level, some countries insist that they are being unfairly punished in performance terms when affected by the trans-boundary effects of pollution (i.e., air pollution that travels from one country to another). The EPI has made the decision to not take these trans-border impacts into consideration; however, it may be possible to account for these effects using modeled data or on a smaller regional scale. YCELP conducted an analysis in 2010 examining the linkages between environmental performance and trade.[15]

▶ Keep in mind the target audience.

As described in Chapter 2, it is important to keep the target audience and users of an index in mind when selecting appropriate indicators. For example, if the key audience of an index is the general public, easily understood indicators based on readily available data may be needed, as opposed to an index that may be used by businesses or an industry association that could entail technical indicators laden with scientific jargon or terms of art. In the latter case, more detailed indicators with specific data may be required to provide final information.

▶ Establishing a set of principles or criteria can help select indicators.

A set of principles, such as those discussed in Chapter 2, can help to determine which indicators to include in an index. *Box 4.1* describes six principles: relevance, performance orientation, established scientific methodology, data quality, time series availability, and completeness that we used for the 2012 EPI and Pilot Trend EPI. We found these criteria helpful when evaluating hundreds of datasets for inclusion in the EPI.

▶ Conduct a cost-benefit analysis to review resource and time constraints for data collection.

It may be possible for developers to collect data, particularly if an index is project-based. It is important to consider the feasibility and the time requirements of data collection. If indicators are not designed in a pragmatic manner, they may become too costly and impractical to include in an index. A simple back of the envelope calculation or cost-benefit analysis can help determine whether it is practical to obtain data necessary for a particular indicator. Some datasets are readily available and in a useful format that will save the core team valuable time. However, other datasets are less transparent (i.e., lack methodological description) or may require aggregation or further research and analysis to be useful for an index. These implications should be reviewed to determine if the value of the dataset is necessary for use as an indicator and will provide substance to an index, or if the time required to prepare the dataset for use is beyond its worth. In the case of a global project such as the EPI, we rely

[15] Yale Center for Environmental Law and Policy. 2011. Exploring Trade and the Environment: An Empirical Examination of Trade Openness and National Environmental Performance. New Haven, CT. Available: http://envirocenter.yale.edu/uploads/publications/Exploring%20Trade%20and%20the%20Environment%2005.24.11%20Web%20Download.pdf.

on international agencies and partners to aid in data collection, because our staff and resource limitations prevent us from collecting the raw data ourselves.

▶ Quality counts.

The accuracy and quality of data used to construct indicators also matters, and processes to evaluate these properties are thoroughly described in Chapter 5. For an index to be credible for target users, the underlying data used to construct the indicators must be carefully considered and evaluated. *Box 5.4* describes an issue with some of the water quality indicators used in the 2010 EPI that were based on inconsistent and incomplete data and ultimately called into question the reliability of the indicator itself.

Conclusion

It is important to build a robust framework by developing a tightly linked system of goals and targets that blend together indicators used to define a set of core categories. A framework is not necessarily set in stone—it can evolve alongside changes in policy priorities, data availability, or aims of an index. The philosophy held by the EPI is that a framework is a continuous work in progress. In the evolution from the ESI to the EPI, we have moved away from a broad scope of indicators related to environmental sustainability to a more focused and smaller set of core indicators. These indicators meet our principles and have associated environmental public health or ecosystem sustainability targets, all of which are necessary elements for helping users decipher the EPI and fulfilling its overall objective.

Box 4.1. Principles used for Indicator Selection in the 2012 EPI and Pilot Trend EPI

The EPI team incorporated six principles to determine which indicators to include in the 2012 EPI and Pilot Trend EPI. Indicators that did not meet these six criteria were not selected for the indicator framework.

Relevance: The indicator tracks the environmental issue in a manner that is applicable to countries under a wide range of circumstances.

Performance orientation: The indicator provides empirical data on ambient conditions or on-the-ground results for the issue of concern, or is a "best available data" proxy for such outcome measures.

Established scientific methodology: The indicator is based on peer reviewed scientific data or data from the United Nations or other institutions charged with data collection.

Data quality: The data represent the best measure available. All potential data sets are reviewed for quality and verifiability. Those that do not meet baseline quality standards are discarded.

Time series availability: The data have been consistently measured across time and there are ongoing efforts to continue consistent measurement in the future.

Completeness: The dataset needs to have adequate global and temporal coverage to be considered for inclusion as an indicator.

Box 4.2. Identifying an Additional Objective for the China EPI

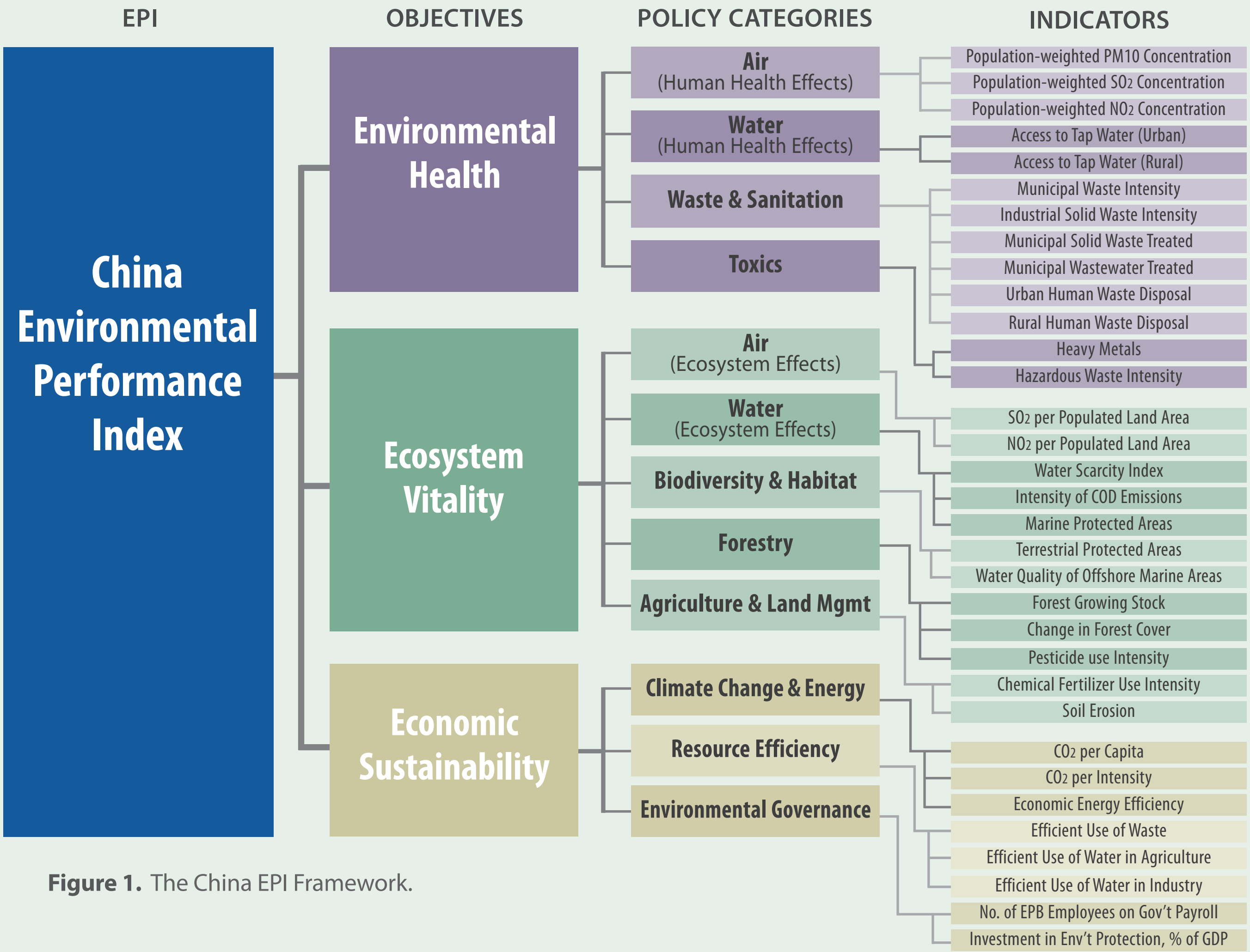


Figure 1. The China EPI Framework.

The indicator framework for the China Environmental Performance Index feasibility study, "Towards a China Environmental Performance Index[16]," was the subject of intense deliberation amongst international and Chinese experts. While the overall structure of the EPI framework, including the Environmental Health and Ecosystem Vitality objectives, remained intact, the inclusion of a third objective on Economic Sustainability was added. The Economic Sustainability objective objective includes indicators primarily focused on the environmental or resource efficiency of economic activities. This is a major policy priority of the Chinese government and was the reason why experts and stakeholders felt these new indicators were critical to include in the final index.

16 Yale Center for Environmental Law and Policy (YCELP), the Center for International Earth Science Information Network (CIESIN), Chinese Academy for Environmental Planning, and City University of Hong Kong. 2011. Towards a China Environmental Performance Index. New Haven, CT. Available: http://www.envirocenter.yale.edu/chinaepi.

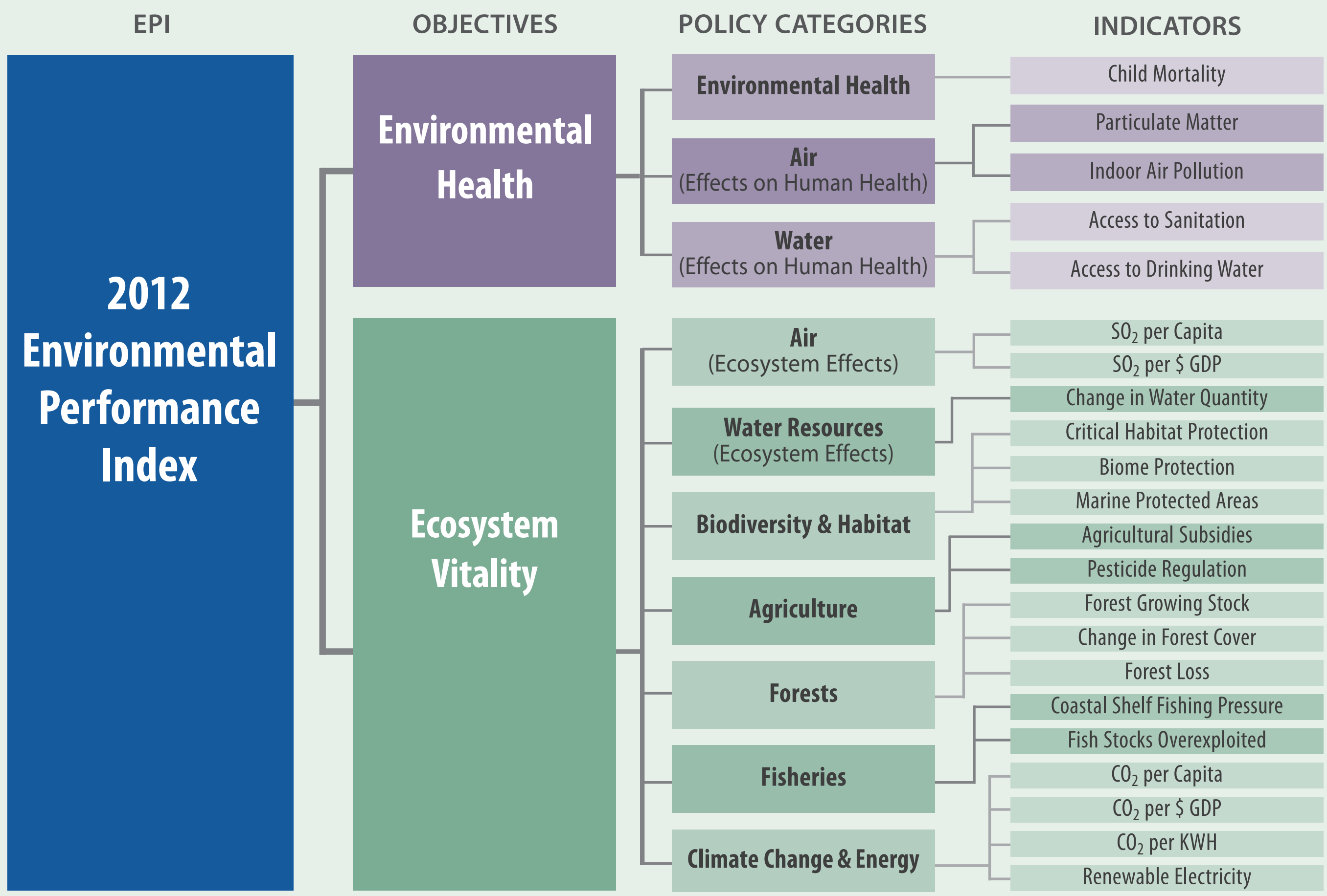


Figure 2. The 2012 Environmental Performance Index Framework.

5 Evaluating Data Quality

Topics Covered in this Chapter:

- **Data quality criteria**
  Establishing standards for data evaluation.
- **Data sources and collection**
  Finding common sources of environmental data that can be used to develop credible indices.
- **Formatting and organizing data**
  Using techniques to organize data effectively.
- **Exploratory data analysis**
  Employing statistical techniques to explore data and spot-check potential problems.
- **Data quality grading**
  Utilizing expert opinion to evaluate the quality of datasets.
- **Time series data analysis**
  Assessing time series and trend data to evaluate the quality of data.

Introduction

Evaluating the underlying quality of the data will aid in the development of a credible index. It is critical to determine if the data are valid (i.e., correspond to the "real world") and reliable (i.e., the results are consistent over time and across observations). There is no single test for validity and reliability, but an investigation into the monitoring systems and adherence to proper protocols are usually sufficient to establish if the data can be used for indicator construction. It is also helpful to run statistical tests for outliers or to compare against datasets measuring the same or similar parameters. If multiple sources of data are available for a given indicator, a thorough vetting of the relative strengths and weaknesses of each data source can uncover anomalies or potential issues with regards to validity and reliability. Availability of time series data allows for examination of trends and patterns with respect to the data, which can also help to reveal anomalies and outliers in a dataset.

This section will explore techniques by which the quality of datasets might be determined. Key considerations for data assessment include: data quality criteria and standards; data sources and collection; formatting and organizing data; exploratory data analysis; expert data quality grading; and time series data analysis.

Data Quality Criteria and Standards

There are several principles by which to determine data quality: quantitative accuracy, representativeness, completeness, and comparability.[17] 'Quantitative accuracy' refers to limits of uncertainty that can be assigned to measured values. 'Representativeness' is the degree to which the data accurately and precisely represent what is trying to be measured, whether it is a parameter, property, characteristic, or condition. 'Completeness' is the comprehensiveness of data coverage, and it is critical in assessing data quality. Incomplete datasets may compromise the decision process for obvious reasons. The last quality indicator is 'comparability' (i.e., having consistent data across units of observation), which is a requirement for comparisons between countries. Other considerations with respect to data quality are further discussed in this chapter.

We adopted our own standards of data quality for the EPI that are distinct from the principles for indicator selection discussed in *Box 4.1* (Chapter 4). The indicator selection criteria help to determine which indicators to include in an index and define the principles of data quality to distinguish between datasets. For the 2012 EPI, we adopted the following principles to help determine which datasets to use:

- **Use of observational data rather than model estimates or imputed data.**
  Observational data is directly measured, while modeled or imputed data uses algorithms to fill in missing data from sparser measurements. Well-developed models can provide accurate and precise estimates of reality, especially where observational data is difficult to find, and can also be used to run scenarios and forecasts. But while modeling and imputation may produce datasets with greater coverage, data accuracy is sometimes compromised.
- **Commitment on the part of the data providers to produce regular updates.**
  Regular updates to a dataset allow for updates and revisions to an index as new data become available. Also, updates allow for temporal consistency between indices—the same data source can be used for the same indicator between editions, ensuring the comparability of index scores across time.
- **Use of datasets with temporal consistency, and avoidance of composite datasets with high temporal inconsistency across countries.** Consistent time series data enable the comparison of results from one year to the next and between countries.

[17] Taylor, J.K. and C. Cihon. 2004. Statistical Techniques for Data Analysis. Boca Raton, FL: CRC Press.

Once a dataset passed these criteria, our team then reevaluated each dataset to determine its appropriateness for the purposes of indicator construction (see Chapter 4 for more information on indicators).

While every attempt was made to find datasets meeting all criteria, in some cases data availability dictated our final indicator selection. For example, to select the most performance-oriented data, our team put data suitability criteria into a hierarchy of importance, then we applied this hierarchy to choose datasets. The first tier of data included measures of direct environmental harm or quality, such as ambient pollution levels to assess air quality. When direct measures were not available, proxy measures (the second tier) were considered as best available substitutes, such as the use of agricultural subsidies[18] to gauge agricultural sustainability under the 2012 EPI's agricultural policy category. Finally, if none of the above tiers of data were available, evaluations of policy intent or motivation were used. An example of this type of indicator is the Persistent Organic Pollutants (POPs) indicator in the 2012 EPI, which was also in the Agriculture policy category. This indicator measures whether countries allow, restrict, or ban one of the 12 "dirty dozen" POPs regulated under the Stockholm Convention.[19] It is not a direct measure of pesticide levels or ecosystem effects but does predict to a certain extent what POPs will likely be present in the environment (i.e., if a POP is not banned there is a greater likelihood that it will be present in the environment). We constructed the POP indicator through a point system to develop an objective measure of policy intent: two points were awarded for a banned substance, while only one point was award for restriction of a POP, and a two-point penalty was applied for any POPs allowed.

Ultimately it is up to the designer of an index to determine what standards and criteria to use for data quality. However, there is an inherent trade-off between employing strict standards of data quality and having data be widely available. An index could implement standards of quality so strict that no appropriate data could meet them. Therefore, while data quality standards can assist with identifying poor quality data, they can also limit the availability of data. For example, it was not until the 2012 edition of the EPI that we incorporated time series data into the index and used a requirement for historical data to

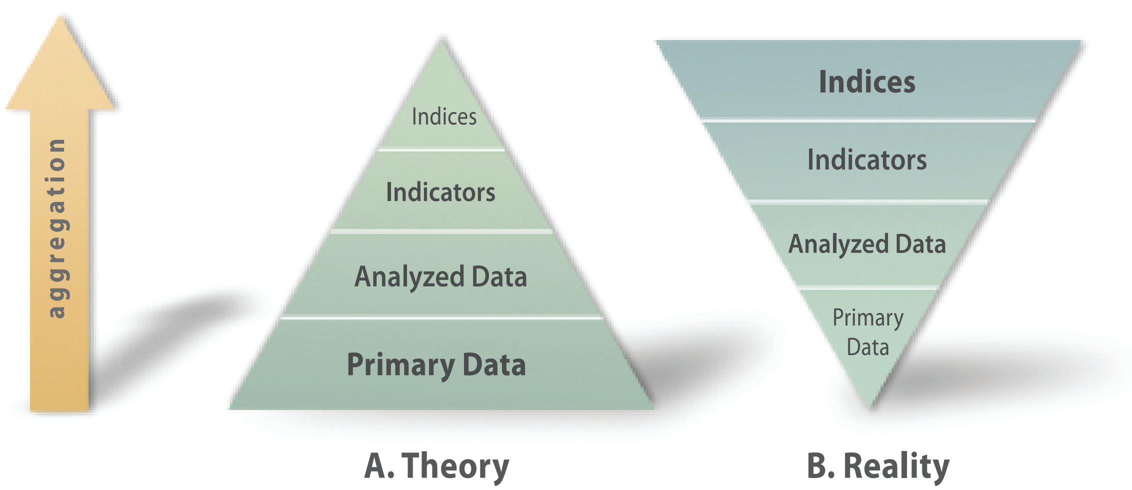


Figure 5.1. Hierarchy of information to create indices (adapted from Segnestam, 2002). In reality, primary data availability is the limiting factor in index development.

18 According to a report by the OECD (2004), public subsidies for agricultural protection and agrochemical inputs exacerbate environmental pressures through the intensification of chemical use, the expansion of land into sensitive areas, and overexploitation of resources.

19 See http://www.pops.int/documents/meetings/inc7/mastlist5/ml5.pdf, page 243 for more information on the specific types of pesticides regulated under the Stockholm Convention.

determine which sources to include. Choosing only time series data limited the number of available and appropriate datasets for the 2012 EPI. The time series criterion also eliminated some otherwise preferable datasets from consideration. For this reason, each data criterion should be carefully considered for its ability to produce to a better index, as well as the limitations it may impose on data selection.

In the end, although the process of data selection for use in an index should resemble the Theory triangle in *Figure 5.1*, with each level of data being selected from a larger pool of data on the lower levels, more often than not the reality is that primary data are often limited, as in the Reality triangle. While index designers would like to have a broad base of primary data available by which to create indicators and ultimately indices, the process of data selection is usually dictated by the design of an index's foundations (Chapter 2), and the data used is determined by whatever data are available. *Figure 5.2* provides a flowchart decision matrix for evaluating a dataset for use in an index.

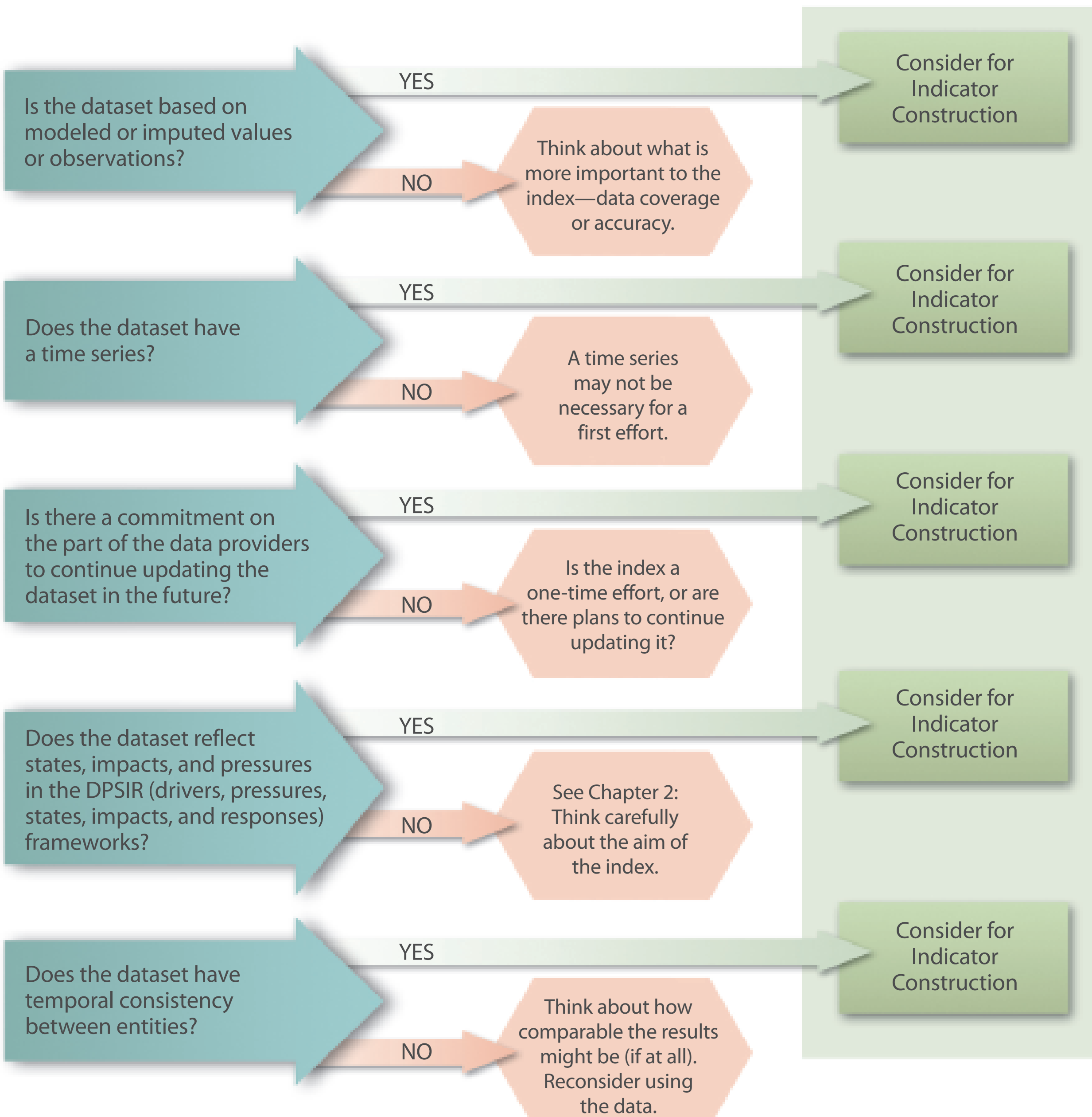


Figure 5.2. Flow chart decision matrix of evaluating a dataset for use in an index.

Data Sources and Collection

The data used for the EPI come from a variety of sources including international organizations, research institutions, government agencies, and academia. There are many sources of data that can be considered for inclusion in an index, such as those from:

- Official statistics that are measured and formally reported by governments to international organizations that may or may not be independently verified;
- Spatial data compiled by research or international organizations;
- Observations from monitoring stations;
- Modeled data;
- Field research and data collection programs through efforts of scientists and organizations (see *Box 5.1* for examples).

Box 5.1. Examples of Field Research & Data Collection Programs

Many data collection programs collect information from researchers, scientists and organizations for different scaled initiatives. Data collected from scientists, members of academia and other organizations are frequently shared through data repositories, or these datasets are incorporated into larger assessments. Below are some examples of how data may be gathered by different sources for sharing or program implementation.

Academia:
Spatial data are often collected by researchers in academia and shared with statewide programs. Universities help provide support by gathering large sources of data that may otherwise be out of reach with a state's resources. The University of Connecticut collaborates with the Connecticut Department of Energy & Environmental Protection to provide maps and geospatial data to the public on the CT Environmental Conditions Online Website, http://www.cteco.uconn.edu. Many of these datasets are provided by and maintained by academic institutions.

Government:
The U.S. government has several large-scale initiatives for collecting national environmental data. The U.S. Geological Survey runs the Cooperative Water Program, which relies on approximately 1,500 state and local agencies to provide data and information on water resources throughout the country, http://www.water.usgs.gov/coop/. The U.S. Geodic Survey collects information for their National Spatial Reference System (NSRS), which establishes reference stations at precisely determined locations (latitude, longitude, height, scale, gravity and orientation), http://www.ngs.noaa.gov. The U.S. Department of Agriculture Natural Resources Conservation provides soil data and information produced by the National Cooperative Soil Survey, http://www.websoilsurvey.nrcs.usda.gov/.

International Organizations:
Some international programs depend upon members of the scientific community and other groups to help collect data and information for global assessments. The data for the IUCN Red List of Threatened Species is collected through assessments performed by members of the IUCN SSC Specialists Groups, appointed Red List Authorities, or by participants of Global Biodiversity Assessment workshops, http://www.iucnredlist.org. They also welcome and review assessments by anyone outside of these member groups. Often, scientists conducting population studies or assessments may contribute their findings to this IUCN program.

Sources for More Information:
U.S. Geological Survey: Cooperative Water Program http://www.water.usgs.gov/coop/
U.S. Geodic Survey: Spatial information (latitude/longitude, etc.) http://www.ngs.noaa.gov/INFO/WhatWeDo.shtml
U.S. Department of Agriculture: National Cooperative Soil Survey http://www.websoilsurvey.nrcs.usda.gov/app/HomePage.htm

Box 5.2. Intergovernmental Organizations (IGOs) that Collect Global Data

We rely on many international data custodians to collect datasets for building the EPI indicators. These data custodians are typically intergovernmental organizations (IGOs) that have the ability to coordinate with countries on a global scale to collect data on a range of issues. Below are several examples of international data custodians, or organizations, many of which we rely on as resources for the EPI.

Food and Agricultural Organization (FAO)
The Food and Agricultural Organization (FAO) of the United Nations collects, analyzes and disseminates data that aids development, including data related to agriculture and agricultural policy. FAO's efforts surround its mandate to achieve food security for all. FAO provides several global spatial databases (http://www.fao.org/corp/statistics) to provide information on topics related to agriculture, such as land-use, water management, production, trade, consumption, nutrition, fisheries, forestry, and imports and exports of wood and paper.

World Health Organization (WHO)
The World Health Organization (WHO) provides leadership on global health matters and is part of the United Nations system. WHO's agenda is focused around its aim to improve public health. To help countries achieve better public health, WHO offers the Global Health Observatory data repository (http://www.who.int/gho)—a collection of data and analyses for monitoring global health. WHO offers numerous datasets on world health statistics, including the occurrence of diseases, rates of mortality, and nutritional information, among many others.

International Energy Agency (IEA)
The International Energy Agency (IEA) provides global information related to its four focus areas: energy security, economic development, environmental awareness, and worldwide engagement. IEA's aim is to help ensure reliable, affordable and clean energy for its member countries and beyond. To help countries achieve energy goals, the IEA provides statistics and online data services (http://www.iea.org/stats), which offer country-level information on important energy related topics, such as emissions, energy statistics and balances, types of energy sources, and renewables.

The World Bank
The World Bank Group consists of several organizations owned by the governments of member nations. The World Bank provides support to developing countries, through policy advice, research and analysis and technical assistance. To ensure countries have access to the best available tools and knowledge, the World Bank provides an extensive database (http://www.data.worldbank.org), which provides information on a range of key topics, from agricultural and rural development, climate change, education, environment, and poverty, to social and urban development.

Sources of data to be used will depend on the size and scale of the issue an index is attempting to measure. The EPI provides a snapshot of environmental performance and only requires the most recent year of data available. Due to the scale of the EPI, which includes over 100 countries, primary data collection is not possible. For other indicator efforts more limited in scope, another option for data collection is through survey or questionnaire techniques, or to allow for entities to self-report data. With the exception of Taiwan,[20] the EPI does not allow countries to self-report data due to potential conflicts of interest and data verification concerns. Instead, for the EPI we often look to organizations, such as those intergovernmental organizations listed in *Box 5.2,* that serve as central agencies for global data collection.

For a high-quality, credible index that allows for self-reported data, we recommend a verification procedure to ensure that data are credible and comparable. *Appendix 4* presents an example of this type of data quality assessment from countries themselves as part of the 2005 ESI. As part of a country data review initiative,

20 For political reasons, Taiwan is unable to report its data to some United Nations-based or other international agencies that do not recognize it as an independent country.

YCELP and CIESIN provided the 2005 data for each indicator to the environmental ministries and statistical offices of 152 countries. We asked the countries to review the collected data for accuracy and provide, where applicable, corrections or recent updates. YCELP and CIESIN also set up a website through which countries could interact with the team as data were being reviewed and updated. A total of 62 countries responded to the data review request, and of these, 25 countries sent updated and additional data. Fourteen countries provided feedback on methodological aspects of the ESI, and 39 countries sent references to reports and websites or informed us that they had no comments on the data sent. This review process ultimately led us to determine that this type of ad-hoc data request was not the best way to obtain consistent country-level data; however, it was a useful exercise in understanding the limitations of this mode of data collection and the value of working with global institutions who serve as centralized data repositories.

When possible, collecting multiple sources of data that attempt to measure the same issue can help to identify potentially problematic data quality issues. Collecting multiple sources of data can provide a basis of comparison between datasets, which can help in determining whether there is a relationship or consistency. If there is little consistency between two datasets, one can investigate why this might be the case and whether one dataset is better suited than the other for measuring the issue of interest. For the 2012 EPI, we evaluated multiple sources of data for carbon dioxide emissions for the climate change policy category, including data from the International Energy Agency (IEA), the United Framework Convention on Climate Change (UNFCCC), and the U.S. Energy Information Agency (U.S. EIA). We created scatterplots of the datasets (one dataset as the x variable, a second dataset as the y variable; see *Figure 5.3*). If the datasets are similar, they should be highly correlated and reveal points close to the identity line of the plot—a 45-degree angle from the origin would indicate high correlation between the two datasets. Any anomalous points in the scatterplot could suggest problems with the accuracy or validity of a given dataset.

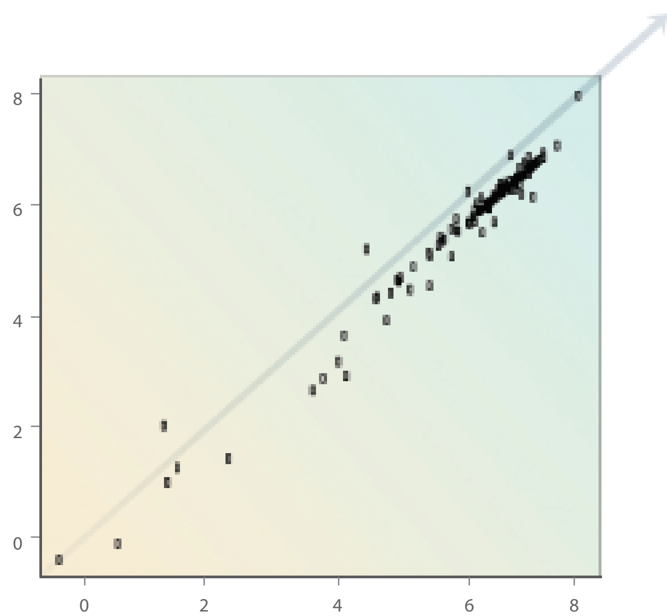


Figure 5.3. Scatterplot of two sources of climate data from the WRI and the IEA respectively.[21] The points fall roughly along the identity line—a 45 degree angle from the origin—and show a strong correlation between the two datasets.

There is another critical rule of thumb when collecting data for the purposes of constructing environmental performance indices: do not throw out any collected data or make any prejudgments as to whether or not a dataset could be useful or pertinent! It is always better to have multiple sources of data and extra data that may come in handy later. Having multiple sources of data also aids in transparency. Often, we are asked why we chose to use one dataset over another. Having additional datasets on hand and detailed notes of our careful evaluation and explanations as to why we chose one over the other helps to build transparency and credibility in the final product.

21 2007 carbon dioxide emissions data is from the International Energy Agency (IEA) on the x axis; 2007 carbon dioxide data is from the World Resources Institute, Climate Analysis Indicators Tool (CAIT) on the y axis.

Formatting and Organizing Data

Consistency is the most important principle to keep in mind when formatting and organizing data for the purposes of analysis and indicator construction. To prevent major confusion in the future, data should be recorded as uniformly and consistently as possible, with the appropriate **metadata** included (i.e., source, author, date of collection, etc.; see *Appendix 3* for an example of metadata from the EPI). Consistent coding of data is also important when there are multiple indicators or datasets used for each issue. Notation of the appropriate units of measure for the data is critical and should be described in detail in the metadata.

For the EPI, we always keep one set of the raw observations and data that come from an original source. This is important if any normalization, transformation, or inversion is applied to a dataset. The original values are maintained in case data issues emerge later in the review. In our EPI metadata, we always include notations with respect to any transformations or changes to the data used for an indicator.

Lastly, data should always be cross-checked. Allow for multiple, independent reviews of data to make sure there are no errors in data entry or formatting. Experts engaged in the development of an index (see Chapter 3) can also help to vet and verify data quality.

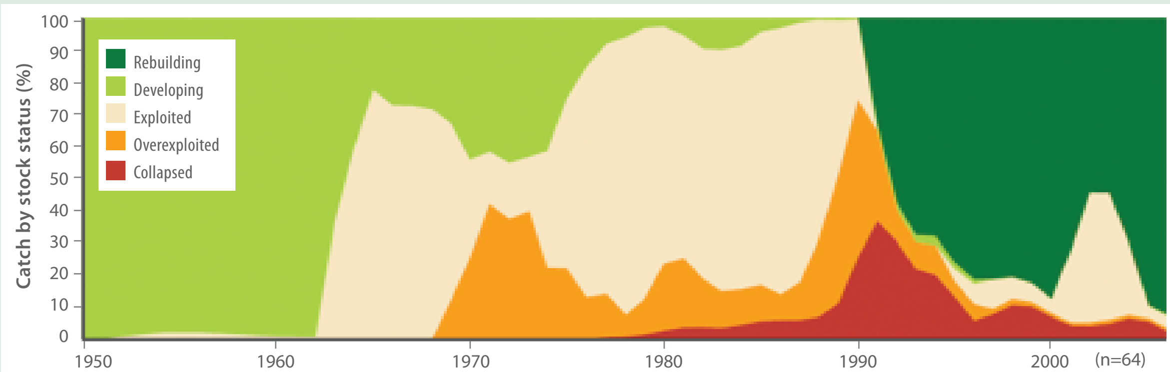
Exploratory Data Analysis

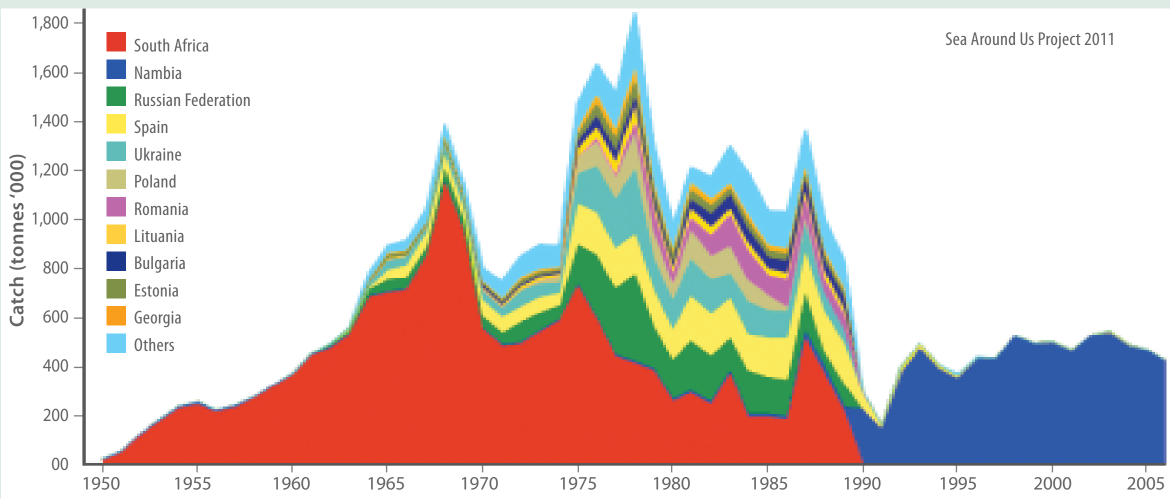
Before attempting to draw any conclusions from more complex analyses, it is important to first examine the data by conducting an exploratory data analysis. Examining the distribution of the data, whether it is normally distributed or skewed left or right, is important for determining if statistical transformations need to be applied. For most datasets, a logarithmic transformation allows more effective differentiation among performers that might be closely clustered together, and a transformation allows comparisons to reflect percentage changes rather than absolute levels of performance (see Chapter 6). Truncation to a range of desired values may be needed to exclude some unwanted values. For example, in the 2012 EPI negative values of the Agricultural Subsidies indicator are truncated to zero because these values actually represent agricultural taxation rather than subsidies, which is what the indicator is intended to measure. In combination, this preliminary data analysis work leads to cleaned and transformed data.

When exploring the data, it is critical to ask whether changes in the data reflect real changes in performance or actual changes in what the proposed index is attempting to measure. What are the sources of the change – could the data reflect mere changes in collection methodology or reporting, or do the data most likely reflect on the ground changes? Often it is difficult to identify the signal through the noise of a dataset to determine whether observed and measurable changes are due to actually implemented policies. *Box 5.3* provides an example of a dataset that shows real change as a result of policy action. One way to evaluate such a change is to compare two similar entities. For example, in the case of the EPI, we can compare the datasets of two countries that are similar with respect to GDP, geography, natural resource endowment, or environmental challenges. If there are differences in an environmental indicator (e.g., forest cover), one can then look at policies in both countries to determine whether the variation in forest cover is due to policy action or simply differences in natural endowments.

Regression analysis and statistical modeling have also been useful in evaluating data quality. Correlation analysis allows for comparisons of datasets from different sources, thereby helping to spot anomalies or inconsistencies in the data. For example, to determine whether an air pollutant emissions dataset such as sulfur dioxide is credible, one could run a correlation analysis of this dataset with energy or coal combustion data, because sulfur dioxide emissions are primarily the result of these processes. If the coal combustion data and sulfur dioxide emissions data do not match or show similar trends, there could be enough basis for questioning the validity of one of these datasets. Often the collection of "ancillary" datasets (i.e., a dataset that is not directly used in the index but can be helpful for comparison, such as the case of coal combustion for air quality data) can be used to develop simple linear regression models or conduct this type of correlation analysis. Applying these statistical techniques can reveal patterns in environmental data and whether the indicator or dataset in question provides the desired signal or measurable objective, particularly in circumstances where calculation methodologies or information on descriptive sources are lacking.

Box 5.3. Detecting Measurable Changes in Fisheries Sustainability in Namibia[22]





Often when evaluating a dataset for measurable changes resulting from government policies, it is difficult to distinguish a signal from all of the noise. Working with experts at the *Sea Around Us* Project based in Vancouver, British Columbia, the EPI team developed indicators for Coastal Shelf Fishing Pressure (the intensity of fishing gear, like trawls, that operate on the shelf) and Fish Stocks Overexploited and Collapsed (the fraction of fish stocks overexploited and collapsed by exclusive economic zone (EEZ)). When trying to interpret the data for these indicators, Namibia is a clear example of where policy interventions have directly led to improvements in fisheries sustainability. These results are even more evident when taking a look at the full time series of data provided by the *Sea Around Us* Project, as demonstrated in the above two plots. In the early 1990s, fish stocks in Namibia were in dire straits, with around 80 percent of stocks in a collapsed or overexploited state (red and orange areas in the upper plot). From about 2000, the situation improves considerably. This improvement tracks the establishment of Namibia's EEZ in 1990, which eliminated foreign fishing fleets (and can be seen in the lower plot in blue). Until Namibia established its EEZ in 1990, South African, Russian, Spanish, and Ukrainian vessels took the bulk of the catch (see figure, top). After 1990, Namibia restricted the access to its EEZ, and was able to enforce restrictions.

[22] Data source: *Sea Around Us* Project, 2011. Fish catches by country from 1950-2005 are shown in the above plot; and Species Stock Status from 1950 to 2005 is depicted in the lower plot.

Data Quality Grading

The Food and Agriculture Organization (FAO) has adopted data quality indicators in their system to assess the quality of their statistics. These indicators include: relevance and completeness, accuracy, timeliness, accessibility and clarity, comparability, and coherence.[23] Our team also conducted a quality assessment of variables used to construct the 2005 ESI. Relying on expert opinion to assess the variables used to construct the ESI indicators, we used a graded quality assessment rubric (see *Appendix 4*) to gauge relevancy (i.e., relativeness between a variable and the issue of interest), accuracy (i.e., reliability of the data source - whether the methodology used to calculate the variable is widely adopted and well established or other data was available to cross check the accuracy of the data), and spatial and temporal coverage (e.g., most recent dataset, frequency of update, spatial coverage, and availability of a consistent time series).[24] Using these criteria, our team, along with outside experts, assessed data variables, giving “grades” of A (Excellent) to F (Unacceptable) and U (Unknown). The final grades were averaged and a single rating given to the variables. As this process was “inescapably subjective” and limited to the knowledge of the working teams and experts, we found this method of judging the quality of data to be problematic, particularly in areas where experts disagreed on the some of the assessment criteria.[15] For the EPI we dropped this exercise, although internally we follow similar procedures.

Time Series Data Analysis

The availability of time series data provides a greater ability for spotting anomalies and erroneous data to improve the quality and credibility of indicators and the overall index. However, for most pilot efforts, the initial use of time series data may not be necessary, depending on the particular aim of the index. While receiving consistent feedback on the value of time series data as a way for countries to evaluate progress and performance over time, we piloted the use of time series data for a few selected indicators in the 2010 EPI (Access to Sanitation, Access to Drinking Water, Industrial Greenhouse Gas Emissions, and Carbon Dioxide Emissions per kWh of Electricity Generation).

For the 2012 EPI, the availability of time series data for almost all of the indicators allowed for more data quality analysis through evaluations of trend data and anomalies. In the case of data for the Fisheries category, for example, we were able to use the time series data to spot erroneous data reporting for some countries that were revealed as anomalies. Experts who provided the data were able to corroborate known reporting issues with the data, which allowed us to remove anomalous data points. In other indicators, for example the Forest Cover and Forest Growth indicators, the availability of time series data allowed us to identify suspicious trends. For some countries, we noticed that data for the Forest Cover and Forest Growth remained unchanged in a five- or 10-year period, causing us to question whether new data were being reported and updated, or if the same data were simply being reported year after year. After examining these patterns, we reviewed the data’s original source and consulted with experts more familiar with the forestry data to determine the origin of reporting issues. We noted these data quality and reporting issues in Section 4 of the 2012 EPI Report.

Time series data can also aid in evaluating whether a particular dataset actually measures or gauges changes in the issue an indicator or index is attempting to measure. With a single data point in time, it is difficult to determine what that data point might be signaling. However, it is important to also keep time

---

23 Food and Agriculture Organization (FAO). 2005. Quality in Statistics: Integrating data quality indicators into the FAO statistical system. Luxembourg: Seventh Eurostat Meeting 23-24 May. http://faostat.fao.org/Portals/_Faostat/documents/data_quality/Integrating_Data_Quality_Indicators_fao_eurostat.pdf

24 Esty, D.C., M.A. Levy, T. Srebotnjak, and A. de Sherbinin. 2005. 2005 Environmental Sustainability Index: Benchmarking National Environmental Stewardship. New Haven: Yale Center for Environmental Law & Policy.

Box 5.4. New Zealand and the 2010 EPI's Water Quality Index

In the 2010 EPI, we included Water Quality Index (WQI) as an indicator in the Water – Ecosystem Effects category. The WQI, which we created, was an attempt to measure countries' water quality in the absence of complete country coverage and data for key water pollutants and measures of water quality. The WQI included three parameters for measuring nutrient levels (Dissolved Oxygen, Total Nitrogen, and Total Phosphorus) and two parameters for measuring water chemistry (pH and Conductivity). We selected these parameters because they cover issues of global relevance (eutrophication, nutrient pollution, acidification, and salinization), and because they are the most consistently reported water quality standards. We collected the data from the United Nations Global Environmental Monitoring System (GEMS) Water Programme, which maintains the only global database of water quality for inland waters, and the European Environment Agency's Waterbase, which has better European coverage than GEMS.

Globally, the UNEP GEMS Water Program is the only international program that collects scientific information on water quality parameters for surface and groundwater, although its future maintenance is uncertain. However, participation is voluntary and each country selects the stations reported to GEMS. As a result, some countries only supply data from one or more monitoring stations; others provide water quality data for almost all of their monitoring stations. Although the GEMS database is the most comprehensive database of water quality, we recognize substantial gaps remain in country, temporal, and parameter coverage, which is why we attempted to create a new measure of global water quality, the WQI.[25]

However, we determined that the data upon which we based imputations and the WQI were not consistent for countries in terms of spatial coverage (i.e. which bodies of water should be included), temporal coverage (i.e. at what frequency), and scientific coverage (i.e., which parameters should be measured). In the case of New Zealand, the country's Prime Minister took the WQI as a standalone measure to claim the country had the "world's second best water quality," behind Iceland.

There are many reasons why this was not a correct assertion. Data were available to compute indicator values for 85 countries, but there were inconsistent time periods for available data. Seventy-four countries had recent data, and 11 countries had data from pre-1990 for which a regression model was used to impute post-1990 scores. A multiple imputation model based on statistical relationships between countries with data and a number of covariates (variables that can predict WQI scores) was used to compute WQIs for an additional 110 countries that had more than 10 sq. km of surface water bodies. Countries with surface water less than 10 sq. km were averaged around.

There was also highly inconsistent station coverage across the countries. For GEMS, some stations are in areas with lower water quality, such as those downstream from pollution activities; others are pristine sites in undisturbed areas or at the headwaters of a water body. Thus some stations measure water that is not representative of overall water quality for an area. One solution to this inconsistency would be to use data from only those stations considered representative. However, based on expert judgment of GEMS personnel we decided not to attempt to identify and filter out non-representative stations. Rather we retained all available data for all countries. We did however apply a station density adjustment factor to penalize countries with fewer reported stations.

Regardless of our attempts, the results were mixed and closer scrutiny revealed that imputed values, rather than ambient or direct measurements, could misrepresent on-the-ground conditions. Therefore, for the 2012 EPI we attempted to shift away from modeled or imputed datasets (with the exception of the Water Quantity indicator) and instead opted for data that adhered to stricter quality standards wherever possible.

25 Srebotnjak, T., G. Carr, A. de Sherbinin, and C. Rickwood. 2011. "A global Water Quality Index and hot-deck imputation of missing data" *Ecological Indicators*, available at http://dx.doi.org/10.1016/j.ecolind.2011.04.023.

scales in mind – a flat trend in a decade's worth of data may not mean that the data are being inaccurately reported or there are issues of quality control. Some environmental changes are slow to manifest themselves in measurable, quantifiable results. As in the case of changes in the global climate, for example a longer time period may be necessary to evaluate a particular change.

Conclusion

For a pilot effort, it may not be necessary to employ very strict data quality standards, particularly if the issue an index is attempting to measure has sparsely available data. Data quality standards represent a process and an evolution. We always release the EPI with a disclaimer that it is a "work in progress" to reflect changes and improvements in data quality and availability over time. The most important rule of thumb is to be transparent about the sources of data, the limitations of data, and any known issues with and information for the data. In this way, one can build a credible index that people can trust and the consequences and limitations of the data used can be easily understood.

With this said, it is still critical to consider the quality of data being used to develop an index. Evaluating data quality is important. For the 2012 EPI, considerations of data quality also helped us determine weightings for the purposes of aggregation (see Chapter 6) for more information on weighting and aggregation). We gave less weight to indicators from datasets with lower quality. Some countries, such as North Korea, were completely excluded based on the quality of the available data. Anomalous results for North Korea on the Access to Water and Sanitation and Fish Stocks Over Exploited and Collapsed (FSOC) indicators raised questions of data quality, leading to our decision to exclude North Korea from the 2012 EPI.

6 Constructing an Index

Topics covered in this chapter:

- **Determining targets**
  Determining high performance and low performance benchmarks.
- **Calculating indicators**
  Calculating performance indicators using a proximity-to-target methodology.
- **Weighting and aggregating the index**
  Selecting and assigning weightings to statistically aggregate the final index.

Introduction

An index typically combines many indicators, as measured by selected datasets, over one or many levels of **aggregation** to produce a single score or rank for each subject. There are many methods for aggregating indicators into indices (see *Appendix 1*), but here we describe the methods used to develop the EPI. In the case of the EPI, indicators are statistically combined or aggregated into policy categories and then aggregated into policy objectives to produce a single score and a global rank for each country with sufficient data. The process for constructing an index involves, at the very least (but is not limited to), the following steps:

- **Determine Targets**
  Before indicators can be calculated, high and low performance benchmarks need to be established.
- **Calculate Indicators**
  Using a **proximity-to-target** methodology, raw datasets are transformed into performance indicators.
- **Weight and Aggregate**
  Weightings are assigned at every level of aggregation, including indicators, policy categories, and objectives according to the relative importance of each level.

In practice, the methodology of indices is far from mechanical. Expert judgment, interpretation of statistical results, and policy or management goals play important roles in setting the many constraints that produce an index. This chapter details both the quantitative and qualitative components for the methodology of index creation based on our experience with the EPI.

Throughout this chapter we use one consistent example (in addition to other examples) to illustrate the methodology described. For this example, consider a car buyer wishing to create a vehicle sustainability index (hereafter called the "VSI") to assist in selecting the most environmentally friendly vehicle based on the combination of four characteristics: vehicle cost, fuel efficiency, passenger capacity and the mining footprint of its manufacturing process. These data will be used to create four indicators: affordability, emissions, capacity and eco- impact.

Determining Targets

The **targets** set for an index establish goals on the indicator level and allow for comparability between entities. Sometimes clear targets exist, either from biological thresholds, policy goals, or from established expert judgment. Regardless, it is the job of the research team to determine targets that best represent the goals of their index.

For the purposes of the EPI, we discovered that the most appropriate available targets for the environmental realm come from international or national policy goals, international organizations (e.g., global organizations such as the World Health Organization, (WHO)), scientific criteria, or expert judgment. The 2012 EPI targets were established using input from five sources:

- treaties or other internationally agreed-upon goals;
- standards or recommendations set by international organizations;
- leading national regulatory requirements;
- expert judgment based on prevailing scientific consensus; and
- ranges of values observed in the data over the duration of the time series.

Selecting appropriate targets is ultimately subjective and relies on the context of an index's objectives. One example to consider is pollution. From an environmental perspective, zero levels of pollution may be

EXAMPLE:

Targets for an imaginary Vehicle Sustainability Index (VSI)

We need four targets for our VSI example, one for each characteristic we have chosen.

Cost: The best cost for a vehicle would be zero, since this represents the lowest price we could pay. However, a new vehicle priced toward the bottom of this scale might raise some red flags, so we will set our target at $10,000 for Cost.

Fuel efficiency: In terms of miles per gallon, higher fuel efficiency is better. The White House has set a fuel economy target to reach 54.5 miles per gallon by 2025. This target falls into the category of "leading national regulatory requirement."

Passenger capacity: A larger vehicle can transport more people per unit of fuel. However, any gains based on passenger capacity are not meaningful above the size of the average group that is traveling to the same location, say, the four people in an average family. A passenger bus may carry more people per unit of fuel, but for the average family, much of that space will often be unused. For this reason, a target above the size of the family would not make much sense. We will set this target at four passengers.

Mining footprint: The "mining footprint" of a car's manufacturing would represent the amount of resources required (in terms of area mined) in order to produce the vehicle. This is measured in hectares, and a car scores better if this value is lower. We will set this target at zero, the least impact possible.

a target value that represents the best possible state of the world. However, a target of zero is far from realistic, because it would mean the cessation of economic activity and a marked decline in living standards. When we chose the target for PM 2.5,[26] we looked to reputable international and scientific organizations to determine what the top performance benchmark should be. According to the WHO, annual average PM 2.5 concentrations of greater than 10 micrograms per cubic meter are known to be injurious to human health.[27] Thus we chose this 10 $\mu g/m^3$ average exposure level as our target for PM 2.5.

In the EPI, achieving or exceeding a target is equivalent to a score of 100 on a 0-100 scale. Along with establishing a high performance benchmark, it is also necessary to establish the low performance benchmark, which is the low end of the indicator range (e.g., equivalent to 0 on a 0-100 scale). For the EPI, the worst-performing country usually establishes the low performance benchmark for a particular indicator, although **Winsorization** at the 95th percentile may also be used to establish this benchmark. For the 2012 EPI and the Pilot Trend EPI, we used the entire time series data (e.g., the lowest performance over a 20 year time series) to set both the low and high performance benchmarks.

There are other considerations when establishing targets. One such consideration may include differential targets, which may be appropriate under certain circumstances (see *Box 6.1*). For example, differential targets may be used if differing levels of economic development make achievement more difficult for a subset of entities included in an index. To set differential targets, one should first determine a characteristic for which differing performance is warranted. For example, data for economic development, such as per capita GDP, should be collected for all observations and then ranked along a continuum with cutoff points

[26] Particulate Matter with a diameter of 2.5 microns or less

[27] World Health Organization (WHO). 2006. The World Health Report 2006: Working Together For Health. Geneva: World Health Organization.

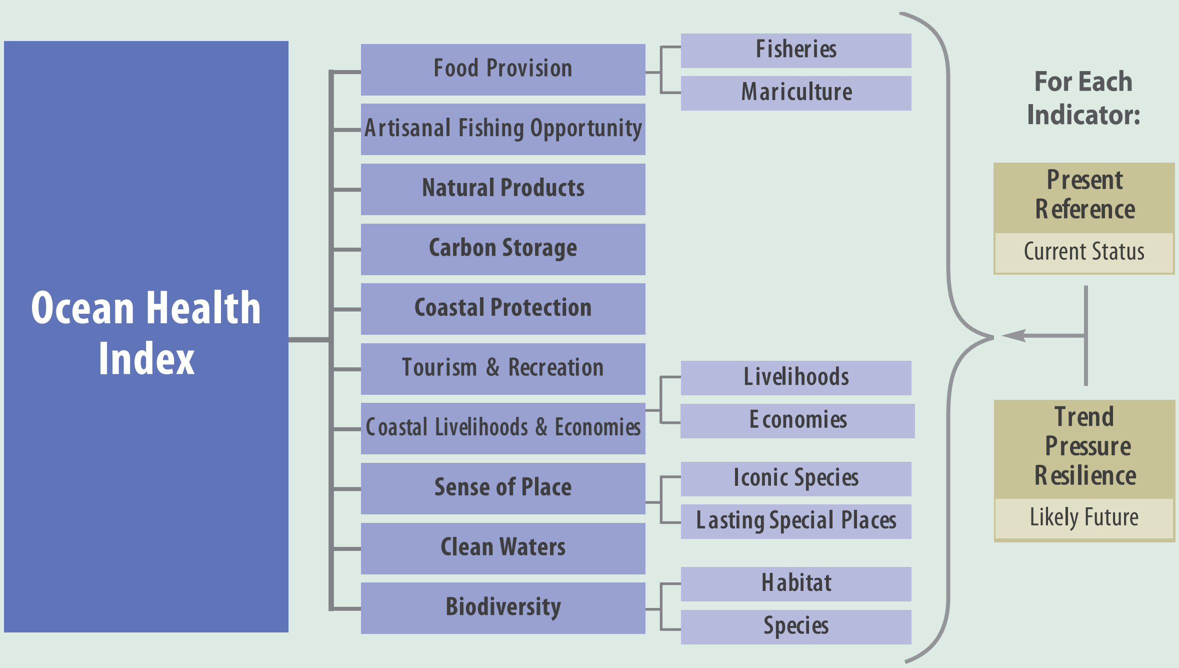
Box 6.1. The Ocean Health Index: Using Differential Targets to Judge Performance

In 2008 there was no definitive measure of the health of the world's oceans. Four years later, using a combination of social, economic and environmental indicators, Ben Halpern, Karen McLeod, Steve Katona, Andrew Rosenberg, Catherine S. Longo and their team have created a multi-dimensional tool to produce just that—the Ocean Health Index (OHI).

One of the unique features of the OHI is that it uses differential targets across countries. This means that the target for good performance is adjusted for each country based on its potential – no country would, for example, be expected to raise the area covered by a given type of desirable ecosystem above that ecosystem's original extent in the country. Through differential targets, the designers of the index measure environmental performance in terms of current status, but also relative to an established reference point. For example, the OHI uses maximum sustainable yield (MSY) to calculate scores for its Fisheries indicator. The fishing landings of each country are judged relative to the country's potential volume of sustainable extraction. Countries that receive the best scores are neither over- nor under-harvesting relative to this value, as it represents the management of the ecosystem for maximum social benefit as well as ecological health. Each indicator in the OHI incorporates a target like maximum sustainable yield that varies across countries. Some are established ecological targets like MSY, while others are historical benchmarks or are based on spatial and functional relationships.

How does the OHI's approach compare to that of the EPI? The Pilot Trend EPI can be compared to the "differential target" approach. In this analysis, each country receives not just an EPI rank and score, but another rank and score based on the trend in its performance for a given indicator over time. In this sense, each country is judged relative to a historical benchmark to calculate its trend score. Both overall score and trend score are presented side by side, such that status and trend can be considered together when forming an opinion on overall performance.

Setting differential targets involves adding another layer of subjectivity to results, as researchers must make judgment calls on how to vary targets across nations. However, it can also elucidate patterns of performance and reward particular types of good performance that would otherwise remain hidden in the data.



for GDP groupings. This might be based on natural breaks, percentiles, standard deviations, or some other measure. Next, a target for each group should be established, for example 100 percent critical habitat protection for the 80th through 99th percentile, 80 percent critical habitat protection for the 60th through 79th percentile, and so on. Based on these new targets, new scores can be calculated that reflect proximity to differential targets, although it must be noted that comparison across groups may now become difficult.

For the EPI, we have foregone differential targets for the reason that comparability then becomes difficult for the entire set of included countries. In the 2012 EPI, we leveled the playing field without having to resort to differential targets by incorporating materiality thresholds (see *Box 6.2*).

Box 6.2. Incorporating Materiality Thresholds in the 2012 EPI to "Level the Playing Field" Among Countries

Recognizing that countries have varying natural resource endowments, physical characteristics, and geography, we applied a concept of "materiality" in the aggregation phase of the 2012 EPI and Pilot Trend EPI. If a country met the criteria for an indicator being "material" (i.e., relevant), we included the indicator in the EPI calculation. For countries that did not meet the materiality threshold, the indicator was "averaged around," meaning the country does not receive a score for that indicator and the weights of the other indicators in a particular category are increased proportionally.

Policy Category	Indicator	Materiality Filter
Biodiversity & Habitat	Marine Protected Areas	Coastal
	Critical Habitat Protection	Must have sites designated as "critical" by the Alliance for Zero Extinction.
Forests	Forest Loss	Must have a minimum 100 sq km of forested land.
	Forest Growing Stock	Must have a minimum 100 sq km of forested land.
	Change in Forest Cover	Must have a minimum 100 sq km of forested land.
Fisheries	Coastal Shelf Fishing Pressure	Coastal
	Fishing Stocks Overexploited and Collapsed	Coastal
Climate Change	Renewble Electricity Generation	Must generate above 130 KWH of electricity annually.

Indicator Calculation

The process of indicator construction includes several steps. In the first step, the raw datasets are cleaned and prepared for use; in particular, missing values and their nature (e.g., country not included in the source data set, country included but value missing, or not applicable) are carefully noted. Second, raw data values (e.g., total carbon emissions) need to be transformed to correct for abnormal distributions. In addition, some data should be transformed by dividing by population, GDP, area or some other denominator in order to make the data comparable across entities (i.e., normalizing the data).

EXAMPLE:

Indicator calculation for an imaginary VSI

For our VSI example, we need to make some transformations to our data. However, no normalizations based on units of observation (population, area, etc.) are necessary in this example because all vehicles are already comparable for our purposes.

Cost: Since the best possible cost is zero, this dataset needs to be inverted so that the best values (low values) receive high value scores. In addition, most measures of money are best viewed on a log scale, in which unit changes can be interpreted as percentages rather than dollars. The difference between a car that costs $5,000 and one that costs $5,500 is about the same as the difference between a car that costs $50,000 and one that costs $55,000. A log scale presents these differences ($500 and $5,000) as the same, since they are the same in percentage terms.

Fuel efficiency: Fuel efficiency is better if it is higher. This dataset doesn't need to be inverted. However, we might consider making transformations if, say the data looks more like Series 1 than Series 2:

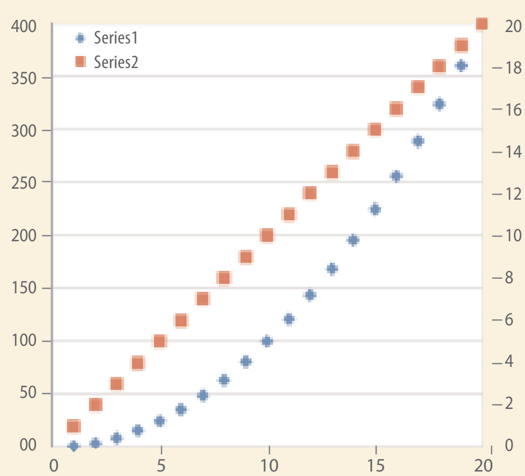


Figure 1. General plot of an imaginary dataset, with Series 2 representing a square-root transformation of the original data (Series 1).

We want consistent units, and a transformation (in this example a square root) would produce the consistency we're looking for.

Passenger capacity: Since our target value for capacity is four, we might consider truncating the data at four. This would produce the same score for vehicles that can carry four passengers as for those that can carry five, since for our purposes the extra person adds no value.

Mining footprint: This data, like the cost data, also needs to be inverted since lower values are better in our index.

The steps for indicator construction and calculation include the following:

1. Data preparation

Common normalizations include percent change (e.g., rates of deforestation over some time period), units per economic output (e.g., energy use per GDP), units per area (e.g., percent territory where water extraction exceeds a certain threshold), or units per population (e.g., CO_2 emissions per capita). Note that the denominator in each case should be relevant for the environmental issue of interest. In some cases, such as air pollution, it may also be useful to weight an indicator by the population exposed. If ambient air pollution is higher in heavily populated urban areas where 75 percent of the population lives, it makes sense for the ambient levels in urban areas to contribute 75 percent to the score for that unit and in rural areas to contribute only 25 percent.

2. Data transformation

In the second step of indicator construction, the **distribution** of the data should be examined to determine if further transformations are necessary. Because the transformed data for the EPI are often heavily skewed, we perform a logarithmic transformation on particular datasets (i.e. datasets that are skewed to the left or right).

Logarithmic transformation serves two purposes. First, and most importantly, if an indicator has a sizeable number of countries very close to the target, a logarithmic scale more clearly differentiates among the best environmental performers. Using raw (untransformed) data ignores small differences among top-performing countries and only acknowledges more substantial differences between leaders and laggards. The use of the **log transformation** has the effect of spreading out performance, allowing the EPI to reflect important differences, not only between the leaders and laggards, but also among the best performers. Secondly, logarithmic transformation improves the interpretation of differences between sub-national units at opposite ends of the scale. As an example, consider two comparisons of particulate matter (PM 10): top-performers Venezuela and Grenada (having PM 10 values of 10.54 μg/m^3 and 20.54 μg/m^3, respectively), and low performers Libya and Kuwait (87.63 μg/m^3 and 97.31 μg/m^3, respectively). Both comparisons involve differences of 10 units on the raw scale (μg/m^3), but they are substantively different. Venezuela is an order of magnitude better than Grenada, while Libya and Kuwait differ by a much smaller amount in terms of percentage on a log scale. Compared to the use of the raw measurement scale, the log scale downplays the differences between the leaders and laggards, while more accurately reflecting the nature of differences at all ranges of performance. This data transformation can encourage continued improvements by the leaders, where even small improvements can be difficult to make, but provides relatively fewer rewards for the same amount of improvement among the laggards.

In some cases, it is also necessary to invert data to make an appropriate fit into an index's framework. This most commonly occurs with the EPI when good performance is on the opposite end of the spectrum from other data. For example, one hundred percent of critical habitat protected implies a high level of environmental performance (i.e., "good" performance), whereas one hundred percent of fisheries overexploited or collapsed implies poor performance. In order to keep high scores on the same end of the performance spectrum, the latter dataset could be inverted by taking the scores and subtracting them from one.

3. Data conversion to indicators

As a final step for indicator construction, the transformed and logged data are converted into indicators to create a common unit of analysis, permit comparability across indicators, and allow aggregation to a composite index using a proximity-to-target methodology, as described below. Different indices use different metrics, such as the ESI's Z-score, the Ecological Footprint's Hectares of biologically productive land, and the Green GDP's use of U.S. dollars. The proximity-to-target methodology measures each entity's performance on any given indicator based on its position within a range established by the lowest performing entity (equivalent to 0 on a 0-100 scale) and the target (equivalent to 100).

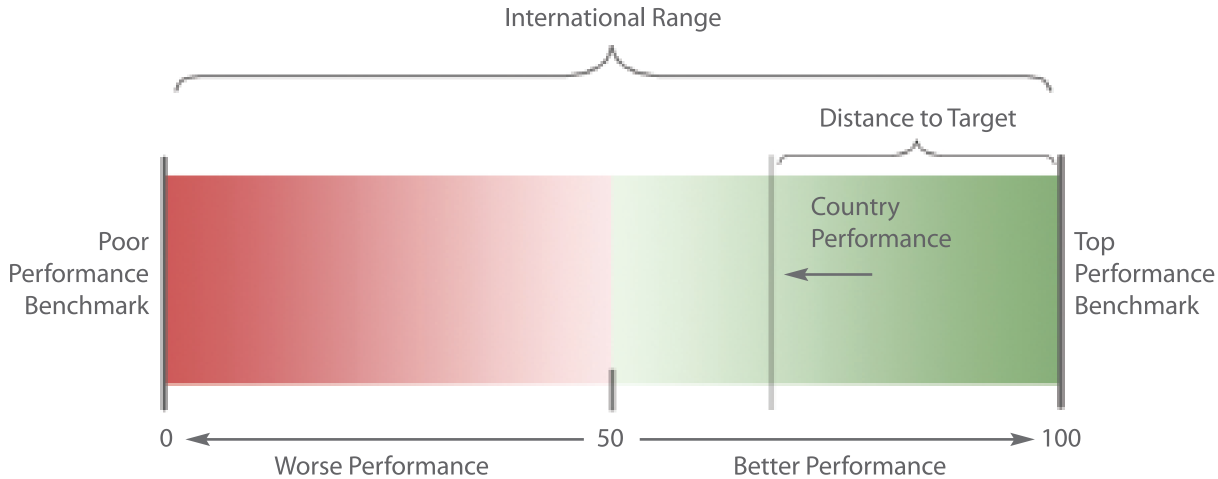


Figure 6.1. Diagram illustrating the proximity-to-target methodology used to calculate performance indicators. Better and worse are relative terms only and refer to distance to the target.

The generic formula for the proximity-to-target indicator calculation in the context of the global EPI is as follows:

$$\frac{\text{International Range} - \text{Distance to Target}}{\text{International Range}} \times 100$$

For example, the score for the Access to Sanitation indicator (i.e., percent of population with access to adequate sanitation) is calculated as follows:

- The target is 100% access to sanitation.
- The worst performer might have 5% of its population with access to adequate sanitation.
- Another country's access to sanitation might be 65%.
- The international range is 100 - 5 = 95.
- For the country with 65% access to sanitation, its proximity-to-target score is calculated as follows: (95 - 35/95) x 100 = 63.1.

EXAMPLE:

Converting data to indicators for the VSI

In order to proceed with our VSI calculations we need all of our data to be in a comparable form. Right now, Cost ranges from 4 to 5 (remember we are using a log scale), Fuel Efficiency ranges from, say, 10 to 50 miles per gallon, Passenger Capacity ranges from 2 to 4, and Mining Footprint from 1 hectare to 100 hectares.

We use the following generalized formula to arrive at our indicator scores:

$$\frac{(\text{Range} - |\text{Target Value} - \text{Actual Value}|)}{\text{Range}} \times 100 = \text{Indicator Score}$$

Here are the conversion calculations for Car A with the following data:

Indicator	Original Data Value	Data Transformation	Transformed Data Value	Original Target	Transformed Target	Indicator Calculation
Cost	$40,000	Log	4.6	$10,000	4	$\frac{(1-\|4-4.6\|)}{1} \times 100 = 40$
Fuel Efficiency	45mpg	None	45	54.5	54.5	$\frac{(54.4-\|54.5-45\|)}{54.4} \times 100 = 83$
Passenger Capacity	2	Truncation	2	4	4	$\frac{(3-\|4-2\|)}{3} \times 100 = 33$
Mining Footprint	80-ha	None	80	0	0	$\frac{(100-\|0-80\|)}{100} \times 100 = 20$

Note that the range for Passenger Capacity is 3 (from 1 to 4) since a vehicle cannot fit fewer than one person, and we have set the upper limit at 4.

We have arrived at our VSI indicator scores for Car A: Cost = 40, Fuel Efficiency = 83, Passenger Capacity = 33 and Mining Footprint = 20.

Weighting and Aggregation

In the field of composite indices, the issues of **weighting** and aggregation are particularly sensitive and subjective. There is no clear consensus among experts on how to best determine a methodological strategy for combining diverse issues, such as those represented in the EPI. Furthermore, the process of assigning weightings is as much of a political process as it is a scientific process. We assign explicit weights to the indicators, policy categories, and objectives that comprise the EPI in order to create the aggregate EPI score (see *Figure 6.2*).

Weights for the EPI were established after considering expert recommendations including perceived quality of data, importance of the indicators and categories for policymaking, and the degree to which the indicators provide direct measurement of environmental performance.

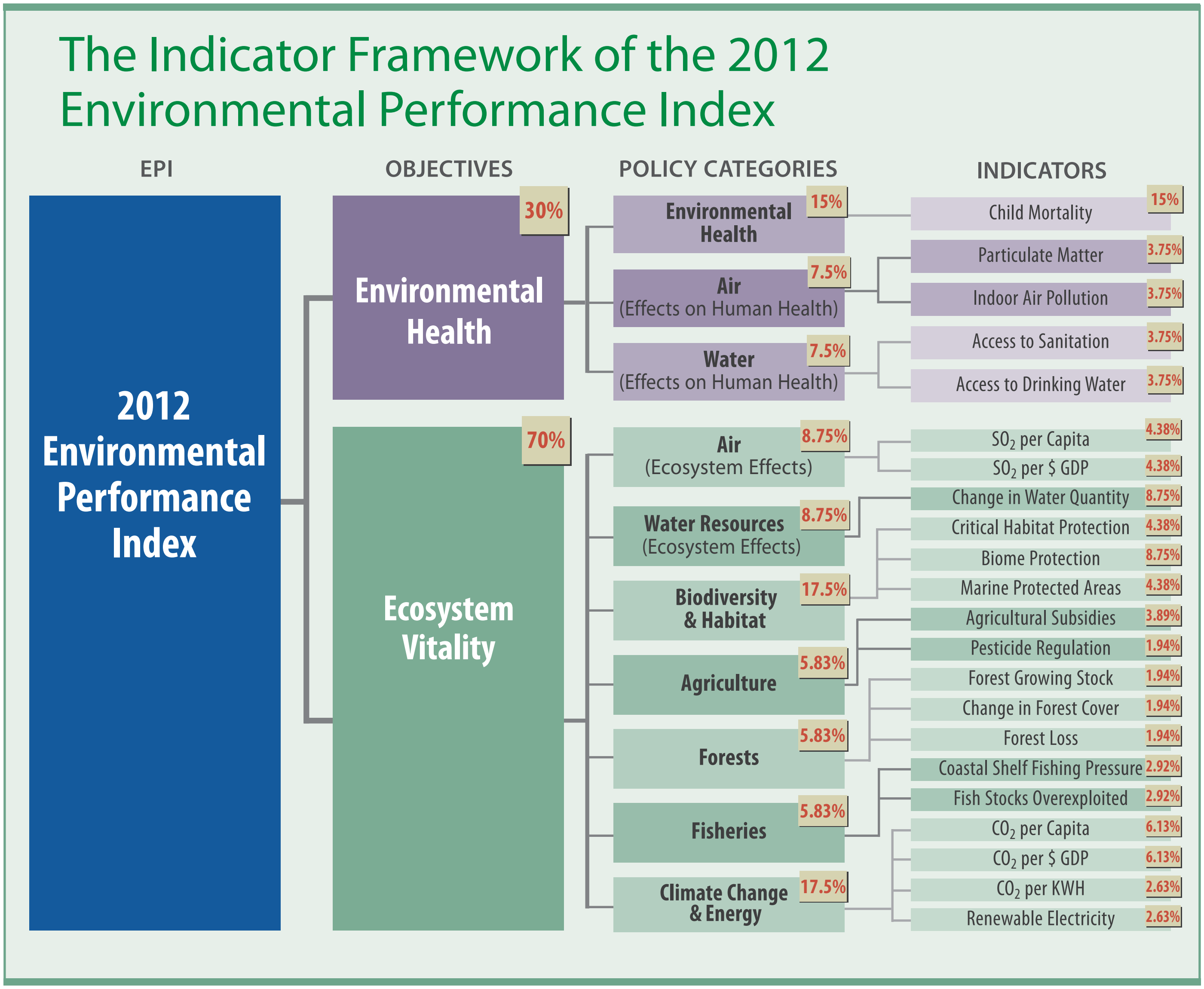


Figure 6.2. The Indicator Framework of the 2012 Environmental Performance Index. The percentages indicate the weightings used for aggregation.

For the EPI, weightings are determined through an iterative process (see *Figure 6.3*). Generally speaking, weightings at every level of aggregation are first divided equally as a starting point. For the EPI, this means the two objectives (Environmental Health and Ecosystem Vitality) initially receive 50-50 weights, each of the 10 policy categories receives 10 percent, and indicators receive weights depending on the number of indicators within a policy category. Weightings can then be adjusted to reflect the relative importance of given issues and the relevance of indicators to category performance. For example, we decided, based on expert input, that CO_2 per Capita and CO_2 per GDP were more important for consideration of climate change contributions than the indicator for the amount of renewable power generation. For this reason, the former two received greater weights relative to the other indicators in their policy category.

Another important consideration when determining weightings is the underlying distribution of the indicator, policy category, and objective scores; or how much variation exists in the data. The impact of underlying score distributions becomes critical when chosen weightings result in composite indices that

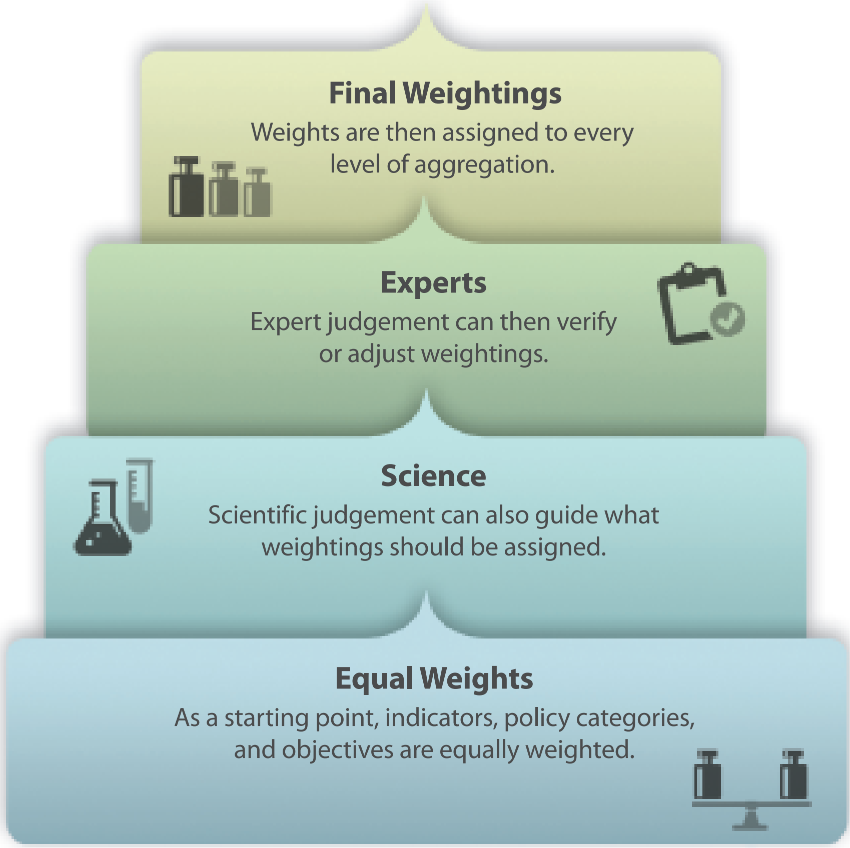


Figure 6.3. A guide to the iterative process by which weightings are determined for an index. This is generally the process used for determining the weightings in the EPI.

do not reflect their initial objective. For example, in 2012 the weights for Environmental Health and Ecosystem Vitality were adjusted to 30-70 to balance the contribution of these indicators to the overall EPI; these explicit weights provide an implicit 50-50 weighting because of the differences in variability of the two policy objectives (less variability across countries is observed for Ecosystem Vitality). When the variability of data for one category is much greater than that of another, it may exert more influence on the overall score if the two categories are given equal weight. What this means is that equal weightings may not always reflect equal influence of objectives or policy categories, which was a consideration for our team when determining the weightings for the 2012 EPI (see *Box 6.3*).

Finally, lower weights can be used to lessen the impact of an indicator for which we have data quality reservations or is more of a proxy for what we really want to measure. In the case of the EPI, we lowered the weighting of the Water (Ecosystem Effects) category because of concerns with respect to the underlying data quality of the indicators.

The EPI team has always advocated for complete transparency in the methodology, recognizing that the selection of weights is not a completely objective process and that disagreements are inevitable based on political preferences and even the performance of individual countries on different facets of environmental performance.

Box 6.3. When are Equal Weights not Equal?

Sensitivity analyses conducted on previous versions of the EPI revealed some unintended results of equally weighting the Environmental Health and Ecosystem Vitality objectives. Despite mathematically equal weights, overall EPI scores were much more highly correlated with Environmental Health than Ecosystem Vitality, revealing the necessity of taking into consideration the distribution of underlying variables when selecting appropriate weights.

For the 2012 EPI, a 50-50 weighting for both the Environmental Health and Ecosystem Vitality objectives meant that the overall composite EPI scores were too heavily influenced by the Environmental Health objective. This unevenness is the result of differences in the variance in the scores for the Environmental Health and Ecosystem Vitality objectives (standard deviations of 27.2 and 12.0, respectively). In other words, the high and low scores on Environmental Health are spread farther apart than those of Ecosystem Vitality. When these more extreme scores are averaged with the more moderate Ecosystem Vitality scores, the former tend to pull the overall score away from the average, and thus produce most of the variation in overall EPI scores. For this reason, countries that perform high in the Environmental Health objective are likely to perform better in the overall EPI, regardless of scores in the Ecosystem Vitality objective.

To correct this statistical imbalance between the two objectives, the Environmental Health objective for the 2012 EPI comprises 30 percent of the overall score while Ecosystem Vitality objective makes up the remaining 70 percent. These relative contributions do not reflect the prioritization of nature indicators over those of environmental health, but rather provide a balance between the contributions of these policy objectives to the overall EPI. They also recognize that humans require healthy ecosystems just as much as they require clean air and potable water. This change in weightings simply reflects a much-needed statistical correction to the aggregation method to produce EPI scores that are more balanced between the two objectives. Environmental Health (EH) and Ecosystem Vitality (EV) have been assigned statistical correlations of 0.57 and 0.64 with the overall EPI score, respectively. The plot below demonstrates the statistical balance achieved through these weights.

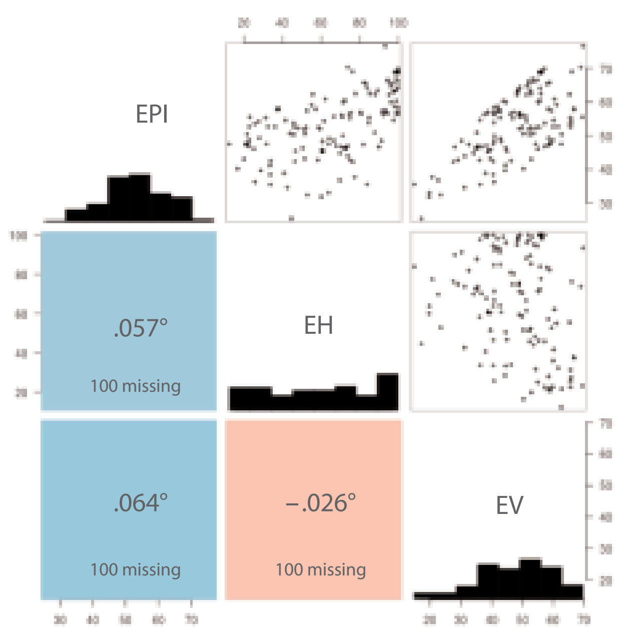


Figure 1. Relationship of the Ecosystem Health (EH) and Ecosystem Vitality (EV) objective scores to the overall EPI Scores.

Other indices can incorporate this lesson by conducting correlation analyses that explore the impact of individual components of the index. A robust index should also include sensitivity analyses that adjust weights and other statistically influential components of the index to uncover places where small changes may have a disproportionate impact on its outcome. Examples of sensitivity analyses are included with each version of the EPI and ESI.

The matrix (left) shows the relationship between Environmental Health (EH) scores, Ecosystem Vitality (EV) scores and the overall EPI. The boxes to the lower left show degree and direction of correlation: .57 for EH-EPI and a higher .67 for EV-EPI. The scatterplots to the top right show the distribution of the data. Environmental Health plotted against EPI has a wider distribution than Ecosystem Vitality plotted against EPI. The relatively widely spaced EH scores tend to pull on EPI scores more than more tightly spaced EV scores, allowing EH to influence more of EPI variation even when the two are equally weighted.

EXAMPLE:

Determining VSI weights

For our VSI let us assume that we value all four indicators equally. We would then be inclined to set their weights all to 25%. However, we should examine the ranges of scores for each indicator before making a final decision:

Vehicle	Cost	Fuel Efficiency	Passenger Capacity	Mining Footprint	Overall Index Score
Car A	40	83	33	20	44
Car B	100	80	66	30	69
Car C	10	75	33	22	35
Indicator range	**90**	**8**	**33**	**10**	**34**

For our three vehicles, the range of Cost scores is by far the largest, and is thus inconsistent with the degree of variation we see in indicator scores in general. Furthermore, we see from the overall index score that the vehicles that scored highest and lowest were also those that scored highest and lowest in the Cost indicator. It appears that the variation in Cost is exerting a greater influence on overall results than that of the other indicators, even though they received nominally equal weights.

To solve this problem, we could adjust the indicator weights according to the range of scores for each indicator:

Vehicle	Cost	Fuel Efficiency	Passenger Capacity	Mining Footprint	Overall Index
Adjusted weight	.04	.46	.11	.39	100
Indicator range	**90**	**8**	**33**	**10**	**34**

This would produce the new index:

Vehicle	Overall Index Score
Car A	51.21
Car B	59.76
Car C	47.11

In this example, the three vehicles are still ranked the same, Car B first, Car A second and Car C third. However, the differences between their overall scores now equally reflect the variation between their individual indicator scores, rather than giving more influence to Cost because of its greater range.

Aggregation Methods

Once indicator weights are established, the individual component scores must be aggregated into an overall score. Most commonly, this is accomplished with a linear aggregation (i.e., adding up the weighted, normalized indicator scores). When weights are equal, this is similar to taking a simple average. However, the major drawback of the linear aggregation approach is that it treats unit changes across indicators and levels of performance as preferentially independent, meaning that the scores do not reflect that a certain level of performance may be achieved in one indicator versus another, such that poor performance in one area can be compensated for by equally (in absolute value) good performance in another area.

Box 6.4. A Work in Progress: Adjusting for a Changing Environment

We always provide a disclaimer that the EPI is a work in progress, to accommodate for updates to datasets, the inclusion of new data, and adjustments to the weightings used for aggregation. One of the benefits of choosing an index framework that encompasses a collection of issues with a transparent and simple-to-understand weighting system is that the index is easy to adjust as prevailing public and scientific opinion changes. As certain issues become more or less important to scientists and stakeholders, the index can be adjusted without starting from scratch.

For example, there was a marked change in public sentiment regarding climate change in the 2008 to 2012 period. For the 2010 EPI, we increased the weighting of the climate change category to 25 percent of the overall score to reflect the overarching importance of climate change to every other policy category in the ecosystem vitality objective. We then equalized the climate change category weighting to be in balance with the other policy categories in the 2012 EPI.

While on the surface this may not seem problematic, it can obscure situations where, for example, an entity performs very well on one or two indicators but relatively poorly on a host of others. With linear aggregation, an entity could score overall as well as, or better than, one that performs relatively well across all indicators, because their average performance is the same. The latter entity may be taking a more holistic approach to environmental governance, while the former may be simply benefitting from a unique case of good management for one or two indicators.

Geometric aggregation is designed to address this issue. In essence, this method punishes extremely low scores more harshly per unit than better scores. Thus, in order to achieve high performance in the above example, a country would have to perform adequately across all indicators and very well in some. A country thus has greater incentive to improve performance in lagging areas, as a one-unit change in performance on a lower scoring indicator would produce a greater improvement in overall score than a one-unit improvement elsewhere. Unfortunately, geometric aggregation is difficult to communicate, and thus reduces the transparency of the index.

But the choice of aggregation methodology depends also on the scientific relationships between indicators, and not just on preferences for trade-offs. Two pollutants in the environment may have drastically unequal effects at different levels, and an increase of just a few units may matter much more with one than with the other. Further, the effects may depend on the units and scales used to measure each (i.e., ambient concentrations, quantities released, etc).[28] For these reasons, there is no absolute basis for comparison between indicators, and choosing an aggregation method should be informed by science but also understood as inevitably a value judgment.

In order to balance these concerns, the EPI uses arithmetic means but addresses different scientific relationships between indicators by using a proximity-to-target methodology, which defines the range of performance with reference to specific high- and low- performance benchmarks. This makes comparison between performance on different indicators possible despite differences in measurement units.

28 See Ebert, U. and H. Welsch (2004). Meaningful environmental indices: a social choice approach. Journal of Environmental Economics and Management, 47(2): 270-283.

EXAMPLE:

Aggregation for the VSI

To simplify our aggregation calculations, let us assume that we retained the nominally equal weights of 25% for each indicator of our VSI.

Vehicle	Cost	Fuel Efficiency	Passenger Capacity	Mining Footprint	Overall Index Score
Car A	40	83	33	20	44
Car B	100	80	66	30	69
Car C	10	75	33	22	35

Linear aggregation adds the individual indicator scores and divides by the total number of indicators, producing the same scores as we discussed previously. Geometric aggregation, on the other hand, multiplies the indicator scores together, then takes the n root of the result, n being the total number of indicators.

Linear aggregation for Car A:

$$\frac{40 + 83 + 33 + 20}{4} = 44$$

Geometric aggregation for Car A:

$$\sqrt[4]{40 \times 83 \times 33 \times 20} = 39.58 = 40 \text{ (rounding)}$$

(Note that these calculations would have to be adjusted if we were using unequal weights)

Vehicle	Overall Score (Linear Aggregation)	Overall Score (Geometric Aggregation)
Car A	44	40
Car B	69	63
Car C	35	27

Whatever weightings and aggregation methodologies are eventually chosen, there will always be individuals who disagree with the final decision. There are several means of giving proper attention to public concern over weightings. The following are examples of how one can address these types of issues, which we have used during our experiences with the EPI.

- **Conduct a sensitivity analysis** that explores how rankings and scores change if the weightings of an index are adjusted.
- **Provide tools** that allow outsiders to weight their own index by adjusting indicator weighting and producing a tailored result based on those weights
- **Supply a clear explanation** as to why weights must be chosen to make aggregation possible, and explain how particular weights represent merely a best attempt at capturing a range of opinions regarding the relative importance of each indicator.

Conclusion

Although this chapter has described a fairly linear process for indicator calculation, weighting, and index aggregation, in reality creating an index is more iterative and may involve multiple revisits to steps in the process. For example, it may be necessary to adjust targets (e.g., adjusting the top performing benchmark to the 97th percentile rather than the 95th percentile) after initial indicator calculations, particularly if the distribution of the indicator scores seems off (e.g., too many entities are at the top performing benchmark). It may also be helpful to involve experts (Chapter 3) during stages in the calculation process to vet the quality and veracity of each indicator and the final index. Our teams work with outside partners such as the Joint Research Centre of the European Commission, which conducts an independent statistical **sensitivity analysis** of the EPI to test the robustness of the calculation methodology and any potential sources of error or bias. In the end, it is essential to properly notate all steps taken in the process of calculating an index in order to transparently communicate how it was developed.

7 Conclusion

Conclusion

We hope that this guide, along with our more than a decade's worth of experience, has provided a useful model by which to develop an environmental performance index. While the process is variable, our best advice is summarized by the following:

- Set the goals, audience and principles of an index to lay the foundation for what issues the index will address, for whom it is created, and what standards to which it will adhere.
- For academic rigor, an index's management team should be complemented by a knowledgeable team of subject area, statistical, and thought-leader experts.
- Balance the best science with the needs of an index's audience to develop a framework, to choose which facets of the issue an index will address, and to reflect the relative importance of those issues in a composite index.
- Indicators should represent the best in data quality balanced by data coverage. Understanding the underlying structure of the data guides adjustments made to the index and produces more appropriate conclusions.
- Based on the knowledge gained through this process, transform data into a format that allows for data to be consistently compared, and select an aggregation methodology to combine indicators in an overall coherent measure.

For more examples of environmental performance indices and how they are being applied in research, civil society, and government for management and policy purposes, we have been compiling case studies that can be found at our Indicators in Practice section of the EPI website (http://www.epi.yale.edu). We hope that these examples can be useful for other organizations seeking to improve their environmental performance through the use of transparent, credible metrics.

In addition, we hope that with the help of this guide we can foster the growth of new data-driven tools for policymakers. In the future, it is our vision that better data will allow for environmental performance indices to be more accurate and to reveal trends between policy action and environmental outcomes.

Glossary

Term	Definition	Chapter
Aggregation	The method by which individual, indicator level scores are used to create a single overall Index score.	6
Audience	The stakeholders and other interested parties who use an index or to whom an index is directed. Defining an audience will help guide the development of goals and principles for an index.	2
Core Categories	A classification of components that reflects facets of its measurable objectives; a method of grouping together related indicators and including an additional level of aggregation. Example: the EPI's indicators are grouped into categories such as Agriculture and Water.	4
Core Team	The (often in-house) research group that will work on the project throughout its life.	3
Data	Raw, quantitative resources that can be developed into information, and then knowledge.	6
Distribution (data)	A pattern of data across a range of values. A "normal" distribution has more data points closer to the mean, and fewer farther away.	6
Environmental Performance	Measurable results of an entity's management of environmental quality.	1
Experts	The (often external) contacts that provide periodic input to guide the project.	3
Framework	The structure, model or design that connects data, indicators, goals, and principles via weighting and aggregation.	2
Goal	The outcome that the index hopes to achieve.	2
Index	An aggregate of environmental performance indicators, which generally implies conversion to common units (or a unitless scale) and weighted or unweighted aggregation (i.e., averaging, adding, or applying other mathematical operations).	1, 2
Indicator (Environmental Performance)	A quantitative variable calculated or measured from an observation (i.e., data) that is used to identify pressures on the environment, environmental conditions (states), or policy responses, or trends thereof.	1, 4, 6
Metadata	Information related to a dataset's source, author, date of collection, etc. that is a necessary component for the data collection process.	5
Methodology	The agreed-upon process by which data collection, calculations, analysis and presentation will be conducted.	6
Metric	A numerical measurement representing data; expressed as an indicator when presented relative to a target.	2
Normalize	To divide one dataset by another as a means of making units unequal and scale comparable; e.g., to arrive at per capita (per person) GDP, total GDP is normalized by population.	2, 6
Objective (Measurable)	The scope of measurable issues related to and provided by the overall objective of an index.	4
Principles	The assumptions that connect methodology and goals; a set of guiding criteria.	2
"Proximity-to-target"	A methodology that measures how close an entity comes to meeting a specified target or goal. Usually the target represents good performance, and "proximity-to-target" methodology involves calculating a score for each entity that shows how well it has performed in relation to that goal.	6
Proxy	A type of data that is used for an indicator as a substitute when a more direct measurement or metric is not available.	4
Scale or Scope	The theme of an index; or the level of index implementation.	2
Sensitivity Analysis	A process by which constraints, such as weights, are adjusted, and then the impact on scores and ranks is observed to determine how sensitive the index is to methodological choices.	6

Glossary (continued)

Term	Definition	Chapter
Smooth	Creating an approximating function that attempts to capture important patterns in the data, while excluding or removing anomalous data or noise.	3
Transformation (Log)	To perform the inverse of taking ten to the power of the number in question. This action transforms data that is scaled by units (e.g., dollars, people, etc.) into data that is scaled by percentages, and often reduces abnormalities in the data's distribution.	6
Target	The measure of "good" performance; the goal/aim/intention that best performers should seek to reach.	6
Weighting	The fraction by which indicators are scaled to determine their relative contribution to overall score.	6
Winsorization	A statistical transformation of extreme (i.e., smallest and lowest, or "outliers") values in data. A given percentage of these values are replaced with observations closest to them and a mean (i.e., Winsorized mean) is calculated.	6

List of Key Acronyms

Abbreviation	Name
CIESIN	Center for International Earth Science Information Network
CPI	Consumer Price Index
DPSIR	Driving Force-Pressure-State-Impact-Response
EGI	Environmental Gender Index
EPI	Environmental Performance Index
ESI	Environmental Sustainability Index
GDP	Gross Domestic Product
HDI	Human Development Index
IUCN	International Union for the Conservation of Nature
KPIs	Key Performance Indicators
MDGs	Millennium Development Goals
OECD	Organisation for Economic Co-operation and Development
OHI	Ocean Health Index
PSR	Pressure-State-Response indicator framework
UN	United Nations
UNDP	United Nations Development Program
UNFCCC	United Nations Framework Convention on Climate Change
WEF	World Economic Forum
WWF	World Wildlife Fund
YCELP	Yale Center for Environmental Law and Policy

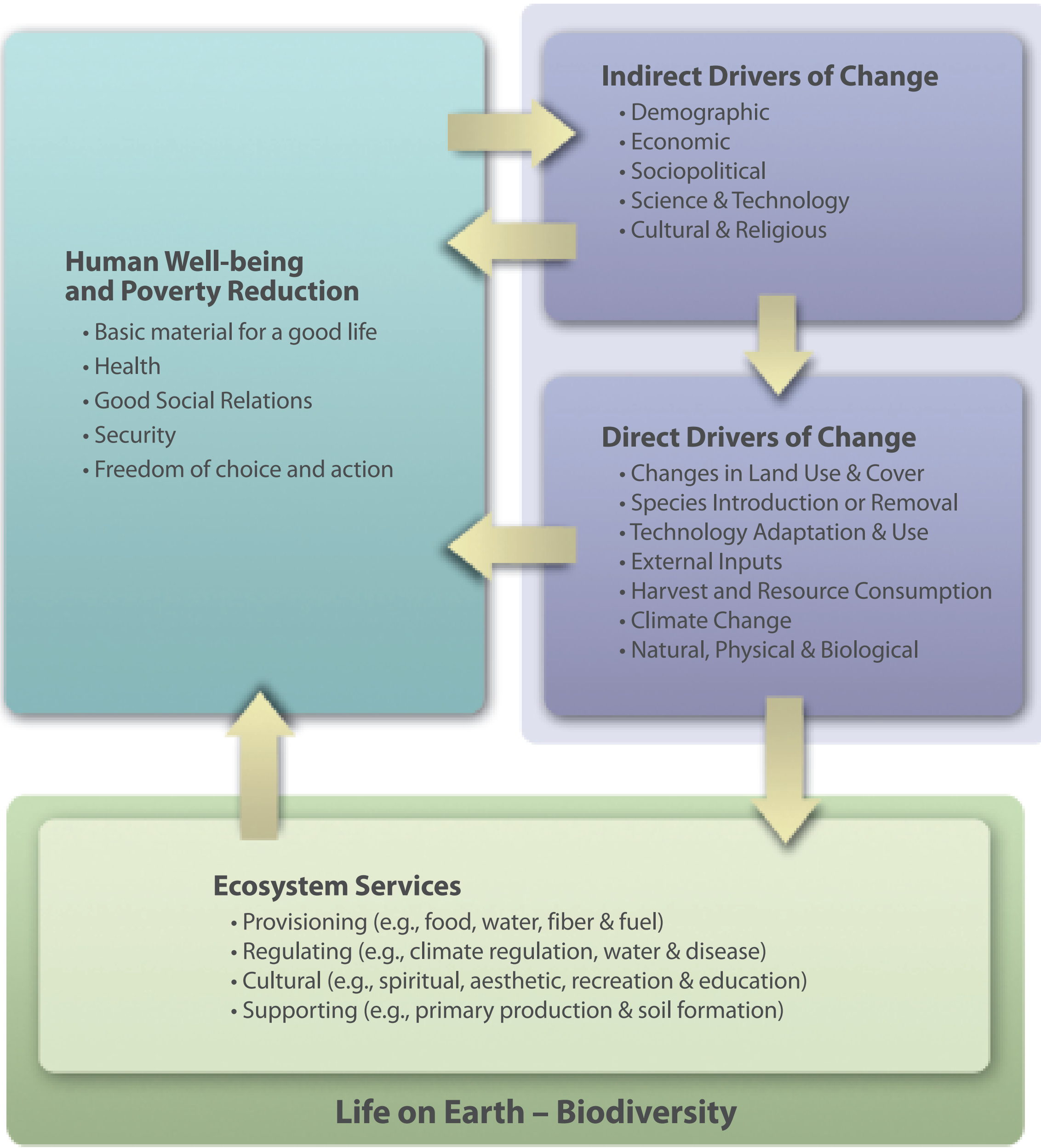
Appendix 1. Additional Resources and Index Efforts

Organization(s)	Year	Resource
OECD	2008	Handbook on Constructing Composite Indicators: Methodology and User Guide
OECD	2007	Composite Indicators: Science or Artifacts?
OECD	2005	What is Sensitivity Analysis
OECD	2006	Sensitive Issues in the Development of Composite Indicators for Policy-Making
UNDP	----	Measuring Human Development: A Primer
UNDP	----	Marketplace for Environmental Training and Online Resources
WRI	----	EarthTrends database

Organization(s)	Year	Index
OECD	----	Competitiveness Index
WWF	----	Ecological Footprint
Reed, Fraser, Morse & Dougill	2005	Integrating Methods for Developing Sustainability Indicators to Facilitate Learning and Action
UNDP	----	Human Development Index
Bell & Morse	2010	The role of Sustainability Indicators within evidence-based policy for sustainable developing in the European Union
Global Footprint Network	----	Ecological Footprint
New Economics Foundation	2011	Measuring our progress - the power of well being
New Economics Foundation	2009	The UN Happy Planet Index (HPI)
Tyndall Centre for Climate Change Research	2004	New indicators of vulnerability and adaptive capacity
NRCS	----	Environmental Quality Index
Ahlheim & Fror - Hohenheim University	2007	Constructing a Preference-oriented Index of Environmental Quality - A welfare theoretical generalization of the concept of environmental indices
UNDP	2008	A survey of composite indices measuring country performance
UN-Habitat	----	City Development Index
IISD	1999	Indicators for Sustainable Development: Theory, Method, Applications
European Commission	----	Sustainable Development Indicators
POINT	2008	Policy Use and Influence of Indicators
South Pacific Applied Geoscience Commission (SOPAC), UN Environment Programme (UNEP)	----	Environmental Vulnerability Index
WWF	1999-2005	Environmental Paper Company Index
Pew Environment Group	2010	Global Aquaculture Performance Index
NASDAQ	2010	Sustainability Index
World Ports Climate Initiative	----	Environmental Ship Index (ESI)
World Economic Forum	----	Sustainability-Adjusted Global Competitiveness Index (GCI)
Hazards and Vulnerability Research Institute	2012	Social Vulnerability Index

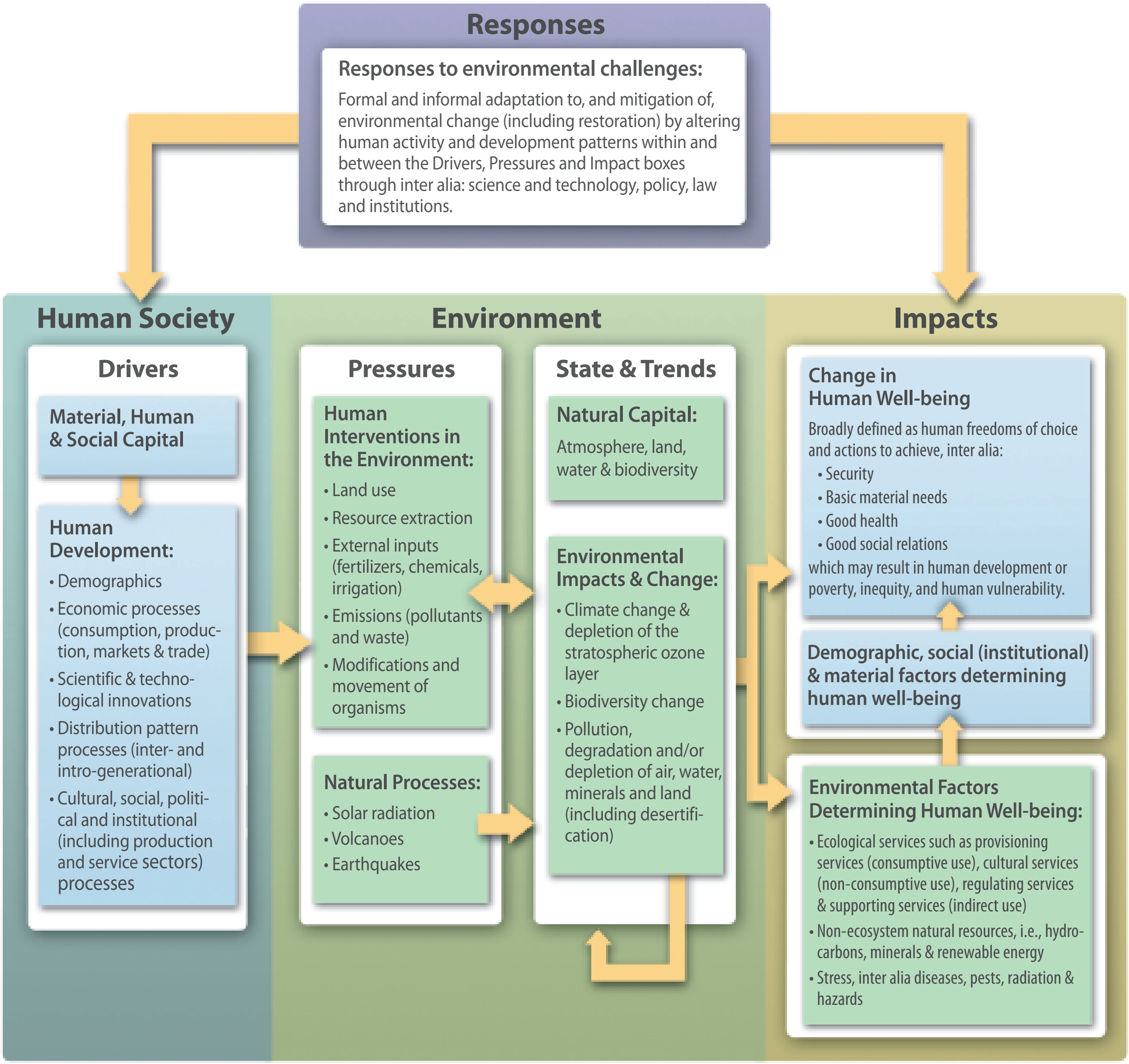
Appendix 2. Examples of Indicator Frameworks

Millennium Ecosystem Assessment Framework



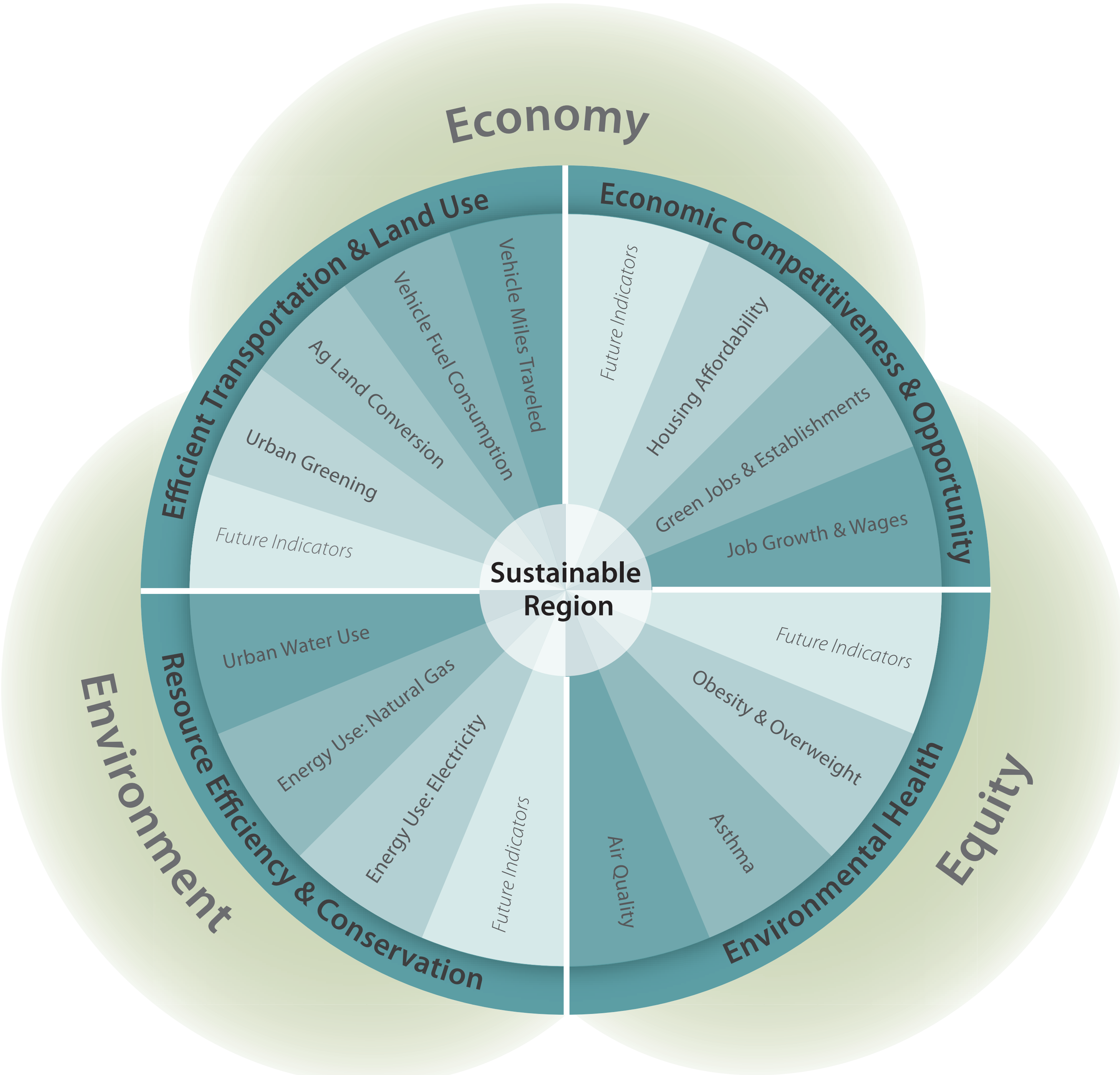
Appendix 2. Examples of Indicator Frameworks (continued)

Global Environmental Outlook (GEO4)



Appendix 2. **Examples of Indicator Frameworks** (continued)

California Forward's **Regional Sustainability Network**



Appendix 3. Example of EPI Metadata

Indicator:	**Critical Habitat Protection**
Objective/Policy:	**Ecosystem Vitality - Biodiversity and Habitat**
Code:	**AZE**
Description:	Percentage of the total AZE site area that is within protected areas.
Rationale:	The Alliance for Zero Extinction (AZE) has identified 587 sites that each represents the last refuge of one or more of the world's most highly threatened 920 species. From the perspective of biodiversity conservation, protection of these sites is of the highest priority.

Source(s)

Variable:	**AZE sites**
Citation:	Alliance for Zero Extinction
Year of publication:	2011
Covered time:	2011
URL:	http://www.zeroextinction.org/
Date data obtained:	10/6/2011
Data type:	GIS polygon shape file obtained from the American Bird Conservancy.

Variable:	**World Database of Protected Areas (WDPA)**
Citation:	UNEP-World Conservation Monitoring Centre
Year of publication:	2011
Covered time:	1990-2011
URL:	http://www.wdpa.org/
Date data obtained:	10/6/2011
Data type:	GIS polygon shape file

Indicator Summary

Unit of Measurement:	Percentage
Indicator creation method:	A time series version of the World Database of Protected Areas (WDPA) from 1990-2011 was obtained from the World Conservation Monitoring Centre. For each country, the percentage area of AZE site(s) that fell within protected areas was calculated.
Additional notes:	The delineation of AZE sites may have uncertainties. Countries with no AZE sites were averaged around for EPI calculations, and are coded -7777.
Transformation needed for aggregation:	none
Target:	100
Low Performance Benchmark:	0
Source:	Expert opinion. The low performance benchmark is the minimum of the 2000-2010 dataset.

Appendix 4. ESI Data Grading Rubic

Component	Indicator Number	Indicator	Variable Number	Variable	Variable Description	Match between variable and issue	Reliability of data source	Variable methodology	Cross-check criteria	Most recent data set	Frequency of update	Spatial coverage	Consistent time series
Environmental Systems	1	Air Quality	1	NO_2	NO_2 Urban population weighted NO_2 concentration	A	A	B	C	A-	A-	D	B-
			2	SO_2	Urban population weighted SO_2 concentration	A	A	B	B-	A-	A-	D	B-
			3	TSP	Urban population weighted TSP concentration	A	A	B	B-	A-	A-	D	C-
			4	INDOOR	Indoor air pollution from solid fuel use	B	B-	C	D+	A	U	B	F
	2	Biodiversity	5	ECORISK	Percentage of country's territory in threatened ecoregions	B	B	C	D	A	A	A	F
			6	PRTBRD	Threatened bird species as percentage of known breeding bird species in each country	B	A-	B+	B-	A	A-	A	A-
			7	PRTMAM	Threatened mammal species as percentage of known mammal species in each country	B	A-	B+	B-	A	A-	A	A-
			8	PRTAMPH	Threatened amphibian species as percentage of known amphibian species in each country	B	A-	A-	B-	A	B	A	B
			9	NBI	National Biodiversity Index	A	A-	B	B	A	U	B	D
	3	Land	10	ANTH10	Percentage of total land area (including inland waters) having very low anthropogenic impact	A-	B	B-	B-	A-	D	A	C-
			11	ANTH40	Percentage of total land area (including inland waters) having very high anthropogenic impact	A-	B-	B-	B-	A-	D	A	C-
	4	Water Quality	12	WQ_DO	Dissolved oxygen concentration	A	B+	B-	B	A	A	D	C+
			13	WQ_EC	Electrical conductivity	A-	B+	B-	B	A	A	D	C+
			14	WQ_PH	Phosphorus concentration	A	B+	B-	B	A	A	D	C+
			15	WQ_SS	Suspended solids	A	B+	B-	B	A	A	D	C+
	5	Water Quantity	16	WATAVL	Freshwater availability per capita	A	B	A-	B	C	A	B	A
			17	GRDAVL	Internal groundwater availability per capita	A-	B	C-	C	A	C	B	D
Reducing Ecosystem Stress	6	Reducing Air Pollution	18	NOXKM	Coal consumption per populated land area	A	A-	B	B+	B	B-	C	C-
			19	SO_2KM	Anthropogenic NOx emissions per populated land area	A	A-	A-	B+	B	B-	C-	C-
			20	VOCKM	Anthropogenic SO_2 emissions per populated land area	A	A-	B	B+	B	B-	C	C-
			21	COALKM	Anthropogenic VOC emissions per populated land area	C	A-	A-	A-	A	A	A	A
			22	CARSKM	Vehicles in use per populated land area	C+	A	A	A-	A	A	A	A-
	7	Reducing Ecosystem Stress	23	FOREST	Annual average forest cover change rate from 1990 to 2000	A-	A-	B-	B	B	B-	B	C-
			24	ACEXC	Acidification exceedance from anthropogenic sulfur deposition	A	C	B	B	D	F	A	F
	8	Reducing Population Pressure	25	GR2050	Percentage change in projected population 2004-2050	A	B	A	A	A	A	A	A
			26	TFR	Total Fertility Rate	A	A-	A	A-	A	A	A	A
	9	Reducing Waste & Consumption Pressures	27	EFPC	Ecological Footprint per capita	A	B	B-	C-	B	A-	B	B
			28	RECYCLE	Waste recycling rates	B	A	C	B-	B	A-	C	D
			29	H[illegible]ZWST	[illegible]neration [illegible]zardous waste	B	A-	B[illegible]	A	B[illegible]	A/B	[illegible]C	[illegible]
		[illegible] W[illegible]	[illegible]	[illegible]AT	[illegible]trial [illegible]ater [illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	A	[illegible]	[illegible]

About YCELP and CIESIN

Yale Center for Environmental Law and Policy (YCELP)

The Yale Center for Environmental Law & Policy (YCELP) was established in 1994 as a joint initiative between the Yale School of Forestry & Environmental Studies and the Yale Law School. YCELP seeks to incorporate fresh thinking, ethical awareness and analytically rigorous decision-making tools into environmental law and policy by providing opportunities to faculty, staff and students to collaborate on projects and research activities surrounding these areas. In addition, the Center supports a wide-range of teaching, research, and outreach initiatives aimed at shaping academic thinking and policymaking. These efforts are also incorporated into the Center's Environmental Protection Clinic and three program areas: Environmental Performance Measurement, Environmental Law and Governance, and Innovation and Environment.

Center for International Earth Science Information Network (CIESIN)

The Center for International Earth Science Information Network (CIESIN) was established in 1989, and in 1998 became part of the Columbia University Earth Institute. CIESIN's mission is to provide access to and enhance the use of information worldwide, advancing understanding of human interactions in the environment and serving the needs of science and public and private decisionmaking. Through the development of several interdisciplinary approaches and data-driven tools, CIESIN supplies valuable information for improving the understanding between human and the environment. In addition, CIESIN supports the research and teaching missions of Columbia University by offering expertise in information technology and data management and contributing to education initiatives.