Practical Challenges in Creating Integrated National-Level Environmental Data Sets:
Lessons Learned from the Environmental Sustainability Index

http://www.ciesin.org/indicators/ESI/pilot_esi.html

Marc Levy, Jessie Cherry,
Alex de Sherbinin, Francesca Pozzi
CIESIN, Columbia University

Abstract

The authors participated in the development of a Pilot Environmental Sustainability Index for the 56 countries included in the World Economic Forum's Global Competitiveness Index. A total of 64 variables went into the index, measuring such attributes as urban air quality, public health, scientific capacity, and the effectiveness of environmental regulation. The purpose was to stimulate dialogue over what constitutes environmental sustainability, how it should be measured, and which policy levers drive it. Problems of data integration are especially acute when working with data sets in diverse formats that were often created for other purposes and at different geographic scales. This poster describes the methods used to overcome these data limitations.

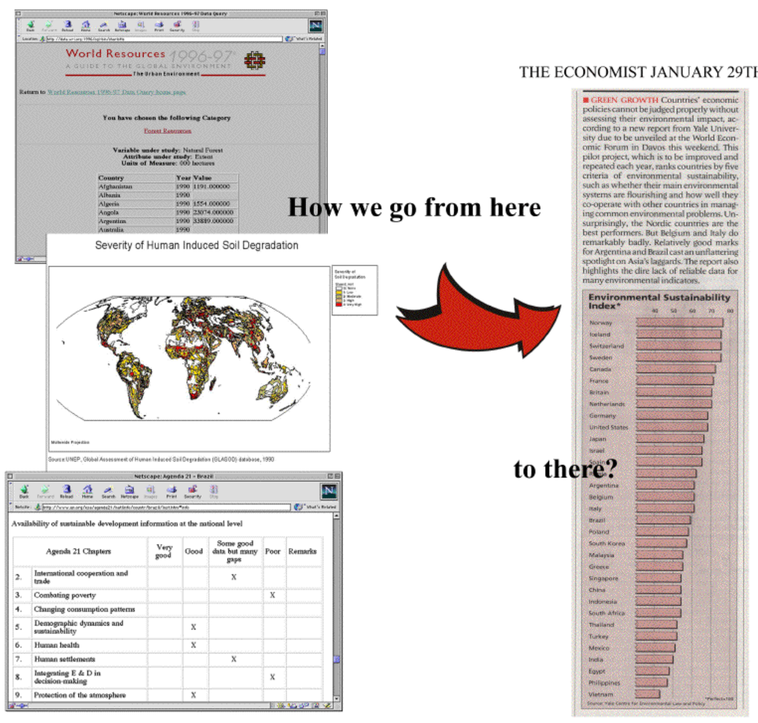
The Challenge

For many purposes, decision makers want information that is integrated into common geographical units and that can be aggregated into a simple, single index. However, the available data are commonly in different and incompatible geographic formats and are not readily aggregated into a single index. The data come in a variety of formats, such as tabular, textual and geographic coverages that may be point-based, raster or vector. Often, data are based on voluntary reporting. In many cases a variable desired on conceptual grounds is simply not available, so appropriate proxies must be identified.

This paper reports on lessons learned through the creation of the Pilot Environmental Sustainability Index, a prototype commissioned by the World Economic Forum Global Leaders for Tomorrow Environment Task Force, in conjunction with the Yale Center for Environmental Law and Policy. A variety of strategies were employed to try to overcome the challenges referred to above, including:

- developing weighting schemes to convert point-based measurements to national aggregates
- using GIS methods to aggregate grid and vector data
- using survey data to augment physical measurements and creating numerical data series from textual reports and other non-quantitative sources
- developing a heuristic structural model of sustainability to permit meaningful integration across various categories of variables.

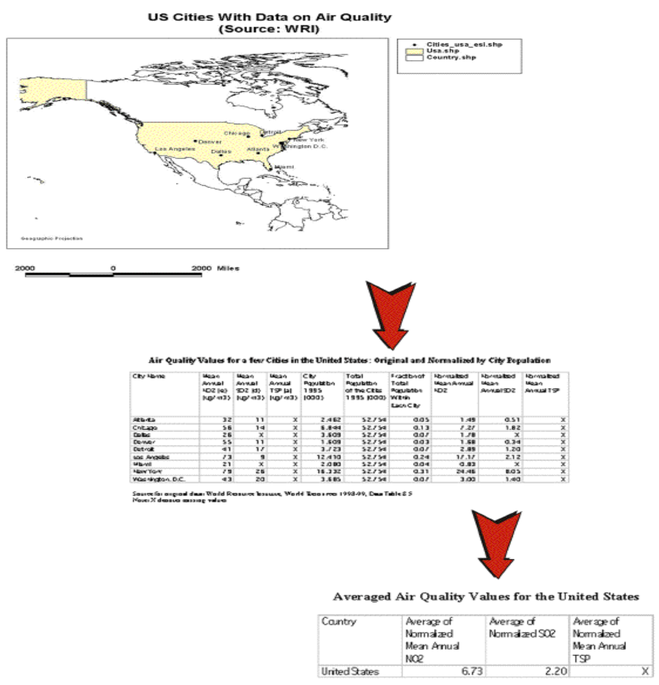
In the development of the Index, it is critical to balance the immediate need for usable policy information with the longer term potential for more accurate and complete data on sustainability.



1. Develop weighting schemes to convert point-based measurements to national aggregates.

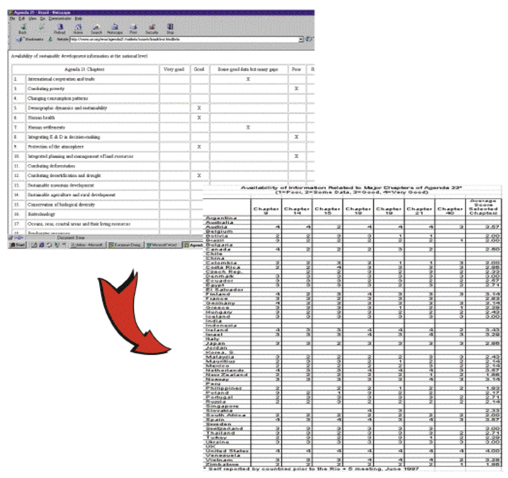
Often environmental information comes from discrete monitoring stations, or applies to fixed points within national boundaries. To combine such data with national-level information, one needs to convert the point values to national-level values through a weighting scheme.

For example, we used urban air quality measurements, and weighted them by the size of the population of the city in which the measurements were taken.



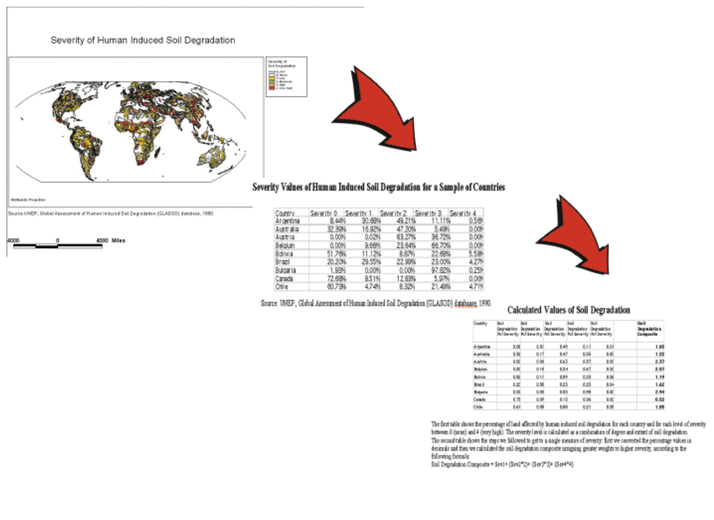
2. Use survey data and other textual information to augment quantitative data series.

Often the desired information is not measured in any quantitative data series, although comparable information is collected through other means. Sometimes such non-quantitative data can be processed to be made compatible with quantitative data sets. Examples used in the Pilot Environmental Sustainability Index include survey data, national reports to international bodies, and membership statistics in international organizations.



3. Use simple GIS methods to aggregate spatial data to national boundaries.

Increasingly, relevant environmental information is available in the form of spatial data, often organized without respect to national boundaries (i.e. grid format). Simple GIS processing can create national-level aggregates to permit these data to be integrated with other national-level data.

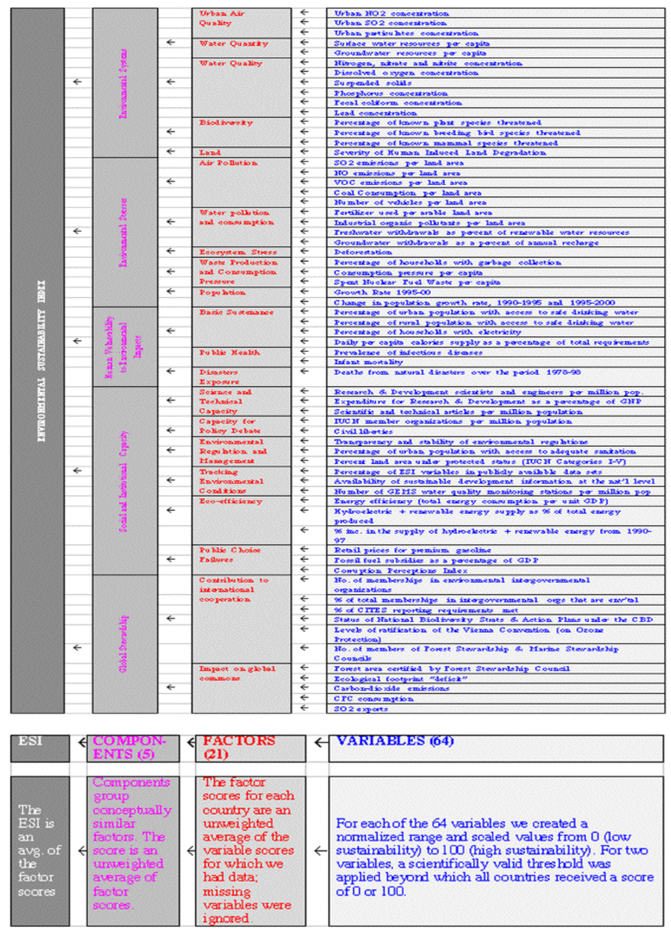


4. Normalizing environmental data by the appropriate geographic area or population.

Data for environmental phenomena such as emissions or hazardous waste disposal are often concentrated in industrialized areas within a country, and may only be of relevance in so far as they affect the people and ecosystems in those areas. Countries such as Australia, Canada, and Russia contain vast tracts of almost uninhabited land. Dividing emissions (e.g., SO_2, NO_x, VOCs) by the total land area yields potentially misleading results. For the Pilot Environmental Sustainability Index, we calculated the total land area with a population density over 5 persons per square kilometer using CIESIN's Gridded Population of the World database.* The resulting "inhabited land area" was used to normalize gross emissions across all countries.

5. Develop heuristic, structural models to permit combination of otherwise incommensurate variables.

For an increasingly relevant category of analytical concepts (environmental sustainability, social stability, political order, human development), the demand for integration across categories of information exceeds the availability of data that are capable of direct aggregation. Typically, the units are incommensurate, and the appropriate weights to assign different variables are difficult to determine. In such cases a simple structural model, accompanied by a simple, transparent aggregation algorithm, helps overcome these difficulties. These techniques permit the creation of prototype models that can serve as the basis for public debate and scientific review, thereby making it possible to create more sophisticated models and algorithms.



Support for this paper was provided by the U.S. National Aeronautics and Space Administration under Contract NAS5-98162. The views expressed here are those of the author(s) and do not necessarily represent the views of CIESIN, Columbia University, or NASA.

*http://www.ciesin.org/datasets/gpw/globldem.doc.html