Macrosystems ecology: big data, big ecology

Patricia A Soranno
Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI

David S Schimel
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA

Ecologists are increasingly confronted by questions that, in one way or another, involve analysis or prediction across vast geographic areas or time periods. There is little doubt that many of the problems facing environmental systems have broad-scale components. These problems range from understanding the spatial distributions of invasive species to discerning how the local ecology of forests interacts with regional fire patterns to influence continental fluxes of carbon. Although ecologists have been successful at answering research questions and developing theory at fine scales, they are now rapidly adding new techniques to their toolkit that facilitate the study of broad-scaled regional processes, and interactions with fine-scaled and global processes. This is where "macrosystems ecology" (MSE) fits in.

The papers in this Special Issue were prepared by participants in the US National Science Foundation's MacroSystems Biology program. A common theme throughout most of these articles is a seemingly simple but challenging topic – data! Specifically, it's the data required to study large, complicated, and highly variable objects typical of macrosystems research. The amount of data involved in MSE research is far beyond that which a single research lab can collect and process. What then are the options available to ecologists for conducting data-intensive research if they clearly cannot collect, process, or analyze it all on their own? At least some ecologists will have to develop the concepts and methodology for studying ecological systems at broad scales; revitalize the culture in which they work to be even more collaborative, open, and interdisciplinary than it already is; and embrace the era of "big data".

To date, ecologists have used any of four strategies for acquiring ecological big data: (1) *Collate existing small but information-rich datasets to create spatially, temporally, and thematically extensive datasets*. This strategy is extremely difficult, is unexpectedly expensive, and can result in datasets with geographic or temporal gaps. (2) *Compile data from remote-sensing platforms that are spatially and often temporally extensive*. This approach is limited by the fact that the variable(s) measured must be drawn from a narrow set of features that can be observed remotely, and which are frequently proxies for the actual quantity of interest. (3) *Link spatially distributed sensor-based observatories or experiments that use common methods*. Such efforts often require complex and expensive instrumentation, and can also have geographic or temporal gaps. (4) *Launch "big science" programs that span continental scales, use standardized methods, and are designed from the outset to address broad-scaled ecological research questions*. These strategies can be costly; require management, computing, and systems engineering skills unfamiliar to most ecologists; and may be subject to spatiotemporal gaps.

So, which strategy should ecologists focus their efforts on to boldly venture into data-intensive research? The answer is all of them. The various approaches for collecting big data have different strengths and weaknesses, and data-intensive science of ecological systems (ie "big ecology") will best progress when all strategies are harnessed to their full potential. Some scientists have dismissed big science approaches in ecology because the International Biology Program (IBP) of the 1960s and 1970s is today frequently portrayed as a failure. However, its legacy may be due for a reassessment: the IBP provided lasting foundational science and datasets used to this day. It can also provide valuable lessons, both positive and negative. Ecologists have made progress over the past 40 years in developing and applying novel methods to address problems across a wide range of scales.

The experiences of the emerging MSE community, some of which are discussed in this Special Issue, demonstrate that ecology needs to integrate the single-investigator model of science with a collaborative, open, and interdisciplinary one. Between the extremes of big science and single-investigator science is a wide range of research conducted by groups of varying size, as small or large teams, working groups, networks, and networks of networks. This Special Issue highlights the growing emphasis on collaboration and a culture shaped by focused, broad-scale scientific questions. Although individual investigator-driven research is still the dominant mode of ecological research, the current successes of MSE research suggest that it is only one of several different possible approaches.

To understand and solve many of today's problems, ecologists need "big data" and "big ecology". This Special Issue of *Frontiers* provides a wealth of new perspectives on this necessity. Hampton *et al.* (*Front Ecol Environ* 2013; **11[3]**: 156–62) asked whether the leaders of big ecology will even be ecologists: this issue suggests the answer is an emphatic "yes".