**EAGER: Managing our expectations: quantifying and characterizing misleading trajectories in ecological processes**

**Overview**

A fundamental problem in ecology is understanding how to scale discoveries: from patterns we observe in the lab or the plot the field or the region, or bridging between short term observations to long term trends and trajectories. In this proposal, we describe a method to directly address the temporal aspects of scaling ecological observations by leveraging existing data produced at more than two dozen Long Term Ecological Research (LTER) sites, an NSF program in place since the early 1980s. LTER sites have produced time series data documenting various ecological phenomena, some going back nearly 40 years. Findings from these sites have been hugely influential in ecology because of their unprecedented longitudinal perspective, yet short-term studies more consistent with typical grant cycles and graduate program are still the norm. We directly address bridging this gap between the short-term and the long-term with an automated approach: in short, we will repeatedly ‘sample’ moving windows of data from existing long-term time series, and analyze these sampled data as if they represented the entire dataset. We will compile typical statistics used to describe the relationship in the sampled data, through repeated samplings, and then use these derived data to gain insights to the questions, *how often are the trends observed in short term data misleading, and can we use characteristics of these trends to predict our likelihood of being misled?*

**Intellectual Merit**

This study will result in the advancement of understanding of how data-driven approaches can be used to foster discovery in ecology, and thus has both data science and disciplinary applications. In ecology, it is widely recognized that understanding long term and broad scale processes are important, but scaling between the fine and the broad remains an open challenge. Similarly, many scientists agree that data archiving has the potential to advance research, particularly in conservation and environmental sciences, where evidence-based decision-making is essential to human well-being. This study will demonstrate a deeply synthetic approach to data reuse with applications in the practical understanding of ecological systems. We will specifically develop paths to data reuse, using best practices; and disciplinary, i.e.: how can we interpret trajectories in ecological systems, given the properties of the system? The disciplinary findings of this study have the potential to be transformative, and of broad influence, because they will provide an empirical framework for interpreting how we understand trends in ecology. The data synthesis aspect will provide a clear demonstration of downstream data use to data producers, and help demonstrate challenges in data reuse and identify best practices to support future synthesis.

**Broader Impacts**

The existing body of ecological data is incredibly valuable for a wide variety of investigations, but largely underutilized by scientists outside the organizations producing these resources. This underutilization likely stems less from a lack of awareness or lack of perception of value, and more from a lack of visible entry points to other scientists. We will provide a framework for downstream use of data produced by others, so that our investigation can meaningfully be used as a starting point for other scientists. Our commitment to ‘radical openness’ of data, intermediate data products, and analysis code not only will ensure reproducibility of the present work, but provide future scientists with both clear entry points, and a model for building their own open work with this communal data resource. Similarly, we will demonstrate the use of open analytical workflows (i.e. reproducible coding, open data, open note-booking) to mainstream academic biologists- particularly as the challenges arise from using data produced by others. In addition to technical documentation, we will also conduct specific outreach on the topics of data reuse. We will use “Practical Data Management (For Bug Counters)” as an existing, established platform for dissemination of accessible, plain language documentation, commentary and engagement around the challenges arising data reuse.