Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-03-22 15:41:38
REFERENCE NO: 189

Author Names & Affiliations

- Christopher Still - Oregon State University
- Daniel Griffith - Oregon State University

Contact Email Address (for NSF use only)

(Hidden)

Research Domain, discipline, and sub-discipline

Ecology, Ecophysiology, Ecosystem Science, Biogeography

Title of Submission

Cyberinfrastructure challenges at the crossroads of evolution and ecosystem sciences

Abstract (maximum ~200 words).

Life plays a fundamental role in the movement of carbon, water, and nutrients globally. Predicting the impact of environmental change on these key human resources requires us to synthesize our understanding of evolution and ecology. We argue that investment in cyberinfrastructure should focus on developing stronger connections among disciplines, as these linkages will provide the largest challenges and the biggest opportunities.

Question 1 Research Challenge(s) (maximum ~1200 words): Describe current or emerging science or engineering research challenge(s), providing context in terms of recent research activities and standing questions in the field.

The biosphere is integral to global carbon, water, and nutrient cycling. Anticipating the impacts of environmental change on the distribution of these key human resources requires that we can predict the responses of biomes, ecosystems, communities, and species across spatial scales. However, facilities for making these predictions are highly limited due to challenges including, for example, how to best represent the functional diversity of life (Peaucelle et al., 2016; Li et al., 2017), how to link disparate data sources (Franklin et al., 2017), and the incorporation of community level processes (Scheiter et al., 2012, 2013). As data synthesis becomes commonplace, the global-scale studies that emerge indicate that disjunct biomes across continents differ in their functioning due to the evolutionary and biogeographic histories of the included species and ecosystems (e.g., Lehmann et al., 2014; Moncrieff et al., 2015; Higgins, 2017).

A major limitation to predicting the future distribution, structure, and function of ecosystems is the availability of a cyberinfrastructure that facilitates integration of data across disciplines. Big data in Ecosystem Science currently consists of several large dataset efforts within different areas of research (e.g., plant trait datasets, or vegetation plot datasets), with some looking to bridge the gaps (e.g., grassportal, BIEN) (Franklin et al., 2017). In most cases, it is advisable that the original scientists that collected the raw data be involved in decision-making and data usage, a reflection of the caveats that come with each dataset within each field (Franklin et al., 2017). Similarly, the development of common metadata formats (e.g., Ecological Metadata Language) (discussed in Jones et al., 2006) has increased the accessibility and transferability of datasets across different fields of study. Despite these advances in Ecoinformatics, the scientific community lacks the necessary cyberinfrastructure to link phylogeny, biogeographic history, species traits, species distributions, community dynamics, feedbacks and disturbance, and environmental layers. Merging these datatypes at a global extent is an ambitious goal, but is required to preserve the biosphere and the functioning of the Earth System. Surely, linking any of these categories and translating this knowledge into models will be transformative for our understanding of the biosphere and the diversity of life.

References
Franklin, J., Serra-Diaz, J.M., Syphard, A.D. & Regan, H.M. (2017) Big data for forecasting the impacts of global change on plant communities: Big data for forecasting vegetation dynamics. Global Ecology and Biogeography, 26, 6–17.
Higgins, S.I. (2017) Ecosystem Assembly: A Mission for Terrestrial Earth System Science. Ecosystems, 20, 69–77.
Jones, M.B., Schildhauer, M.P., Reichman, O.J. & Bowers, S. (2006) The New Bioinformatics: Integrating Ecological Data from the Gene to the Biosphere. Annual Review of Ecology, Evolution, and Systematics, 37, 519–544.
Lehmann, C.E.R., Anderson, T.M., Sankaran, M., Higgins, S.I., Archibald, S., Hoffmann, W.A., Hanan, N.P., Williams, R.J., Fensham, R.J., Felfili, J., Hutley, L.B., Ratnam, J., San Jose, J., Montes, R., Franklin, D., Russell-Smith, J., Ryan, C.M., Durigan, G., Hiernaux, P., Haidar, R., Bowman, D.M.J.S. & Bond, W.J. (2014) Savanna Vegetation-Fire-Climate Relationships Differ Among Continents. Science, 343, 548–552.
Li, D., Ives, A.R. & Waller, D.M. (2017) Can functional traits account for phylogenetic signal in community composition? New Phytologist.
Moncrieff, G.R., Hickler, T. & Higgins, S.I. (2015) Intercontinental divergence in the climate envelope of major plant biomes: Intercontinental biome divergence. Global Ecology and Biogeography, 24, 324–334.
Peaucelle, M., Bellassen, V., Ciais, P., Peñuelas, J. & Viovy, N. (2016) A new approach to optimal discretization of plant functional types in a process-based ecosystem model with forest management: a case study for temperate conifers: Defining the optimal number of PFTs using a trait-based approach. Global Ecology and Biogeography.
Scheiter, S., Higgins, S.I., Osborne, C.P., Bradshaw, C., Lunt, D., Ripley, B.S., Taylor, L.L. & Beerling, D.J. (2012) Fire and fire-adapted vegetation promoted C4 expansion in the late Miocene. New Phytologist, 195, 653–666.
Scheiter, S., Langan, L. & Higgins, S.I. (2013) Next-generation dynamic global vegetation models: learning from community ecology. New Phytologist, 198, 957–969.

Question 2 Cyberinfrastructure Needed to Address the Research Challenge(s) (maximum ~1200 words): Describe any limitations or absence of existing cyberinfrastructure, and/or specific technical advancements in cyberinfrastructure (e.g. advanced computing, data infrastructure, software infrastructure, applications, networking, cybersecurity), that must be addressed to accomplish the identified research challenge(s).

A global synthesis of data linking remote sensing, species occurrences and abundance, species traits, community ecology, biogeography, and evolutionary history is an overarching and lofty goal for Ecosystem Science. A major focus of current efforts is to assemble datasets from the literature in each of these fields. For example, there are many national and international vegetation plot datasets seeking to organize data from individual researchers into broader networks and there are global and regional networks of herbarium specimens. Genbank is another example of big data success within one datatype. However, creating cross-discipline linkages among these groups of datasets will require investment in facilities at the intersection of these fields, especially evolution.

Likely, the most important opportunity lies in organizing synthesis teams and creating cyberinfrastructure that links them to the broader scientific community. The success of organizations such as NCEAS (National Center for Ecological Analysis and Synthesis) and NESCent (National Evolutionary Synthesis Center) indicate that this is a fruitful path for synergy. However, data availability in general and the challenges of merging data both within and among disciplines remains limiting.

We suggest that a parallel, and supplementary avenue for building upon these successes would be to also invest in the collection of a “backbone” of data that directly link multiple disciplines, but do so across a global network (Borer et al., 2014). These networks would contribute high-quality data, systematically collected with the same techniques, spanning key disciplines (e.g., plant traits linked to plant abundance and remote sensing). In addition to providing robust data and metadata to anchor syntheses, these networks would also provide key points for the dissemination of training in Ecoinformatics and the use of HPC and software infrastructure. This network could also identify and address major biases that limit global syntheses, including sampling of underrepresented parts of the world (e.g., Africa) and sampling dominant lineages of plants that are underrepresented in databases (e.g., grasses).

Increased access to computing resources, networks, data storage, and visualization/analysis tools will be important to realize these major goals. However, the major limiting advancements for Ecosystem Sciences are in building cyberinfrastructure among fields of study. The scale of this endeavor will also amplify existing challenges within field related to inclusion of error estimates and documentation of data processing steps when assimilating separate data sources. We argue that funding data assimilation efforts that explicitly link multiple disciplines will be the most successful.

References
Borer, E.T., Harpole, W.S., Adler, P.B., Lind, E.M., Orrock, J.L., Seabloom, E.W. & Smith, M.D. (2014) Finding generality in ecology: a model for globally distributed experiments. Methods in Ecology and Evolution, 5, 65–73.

Consent Statement