

## Finding generality in ecology: a model for globally distributed experiments

Elizabeth T. Borer<sup>1\*</sup>, W. Stanley Harpole<sup>2</sup>, Peter B. Adler<sup>3</sup>, Eric M. Lind<sup>1</sup>, John L. Orrock<sup>4</sup>, Eric W. Seabloom<sup>1</sup> and Melinda D. Smith<sup>5</sup>

<sup>1</sup>Ecology, Evolution, and Behavior, University of Minnesota, Saint Paul, MN 55018, USA; <sup>2</sup>Ecology and Evolutionary Biology, Iowa State University, Ames, IA 50011, USA; <sup>3</sup>Department of Wildland Resources and the Ecology Center, Utah State University, Logan, UT 84322, USA; <sup>4</sup>Department of Zoology, University of Wisconsin, Madison, WI 53704, USA; and <sup>5</sup>Ecology and Evolutionary Biology, Colorado State University, Fort Collins, CO 80523, USA

### Summary

1. Advancing the field of ecology relies on understanding generalities and developing theories based on empirical and functional relationships that integrate across organismal to global spatial scales and span temporal scales. Significant advances in predicting responses of ecological communities to globally extensive anthropogenic perturbations, for example, require understanding the role of environmental context in determining outcomes, which in turn requires standardized experiments across sites and regions. Distributed collaborative experiments can lead to high-impact advances that would otherwise be unachievable.

2. Here, we provide specific advice and considerations relevant to researchers interested in employing this emerging approach using as a case study our experience developing and running the Nutrient Network, a globally distributed experimental network (currently > 75 sites in 17 countries) that arose from a grassroots, cooperative research effort.

3. We clarify the design, goals and function of the Nutrient Network as a model to empower others in the scientific community to employ distributed experiments to advance our predictive understanding of global-scale ecological trends and responses.

4. Our experiences to date demonstrate that globally distributed experimental science need not be prohibitively expensive or time-consuming on a *per capita* basis and is not limited to senior scientists or countries where science is well funded. While distributed experiments are not a panacea for understanding ecological systems, they can substantially complement existing approaches.

**Key-words:** collaborative experimental network, distributed experiment, meta-analysis, metadata, Nutrient Network, relational data base

### Introduction

Integrating research across organismal to global scales presents a major challenge for ecology, because many pressing questions require an understanding of whole-Earth, global-scale processes, whereas species diversity, species interactions and ecosystem function – and often the individual human decisions that alter ecological systems – generally occur at local or regional scales (Levin 1992; Denny & Benedetti-Cecchi 2012). To capture the dynamics of species and their services, ecologists must study and manipulate ecological systems at the scale of organisms. Yet a general, predictive understanding of ecological systems requires knowledge of the degree to which functional relationships measured at one site (e.g. between the effect of nitrogen deposition and species diversity) predict these relationships at other sites and in other environments. The long tradition of single-site experiments in ecology has provided little ability to understand the roles of environmental context

and contingency (i.e. functional relationships that occur only under certain conditions) that are critical for predicting responses of ecological systems to global-scale perturbations. Thus, to address the grand challenges facing the biosphere (NAS 2001), we must develop our ability to use information from local studies to make effective predictions about factors affecting biological systems across many spatial and temporal scales (Levin 1992; Denny & Benedetti-Cecchi 2012).

Ecologists employ a diversity of tools (Table 1) to reveal general functional relationships linking local-scale drivers of ecological diversity and function with regional- and global-scale outcomes. Single-site observational and experimental studies, predictive models and meta-analyses are widely used, and observational networks collecting standardized data (e.g. National Ecological Observatory Network (neoninc.org), Global Lake Ecological Observatory Network (gleon.org), National Phenology Network (usanpn.org)) are growing in number. Each of these tools has generated important data and powerful insights into the factors controlling diversity and ecosystem function and serves an important role in advancing our science. However, each of these tools has drawbacks that limit

\*Correspondence author. E-mail: borer@umn.edu

**Table 1.** Ecological tools commonly used for understanding the factors affecting biological diversity and ecosystem structure and functioning. Each has differing strengths and drawbacks for understanding generality and contingencies of functional relationships in ecology

	Consistent methodology	Causal inference	Realistic complexity	Environmental gradients	Site-specific design
Single-site experiments	✓	✓	?	?	✓
Observational networks	✓		✓	✓	
Process-based models	✓	✓		✓	✓
Empirical/statistical models	✓		?	✓	✓
Meta-analyses		?	✓	✓	
Distributed experiments	✓	✓	✓	✓	

their ability to reveal general functional relationships spanning many spatial scales. For example, inference from meta-analyses is inherently hampered by incompatible methodology and design of existing experiments (Gurevitch & Mengersen 2010). On the other hand, although observational networks are the source of critical environmental data, even when observational networks employ identical sampling protocols among sites, functional relationships are difficult to discern in the absence of experimentation (Hewitt *et al.* 2007). Collaborative experimental networks remain rare in ecology, although they hold great promise for detecting general functional relationships and providing predictive ability of functional relationships that are contingent on local conditions.

Collaborative experimental networks, such as ITEX (Elmendorf *et al.* 2012), LiNX (Mulholland *et al.* 2002), LI-DET (Harmon *et al.* 2009) and Biodepth (Hector *et al.* 1999), help address these issues by providing the context to implement consistent and standardized local designs that act in concert, in a distributed fashion, to examine much larger scale ecological questions that otherwise could not be addressed (Fraser *et al.* 2012). In contrast to local or network-scale observations that rely on natural variability to perturb ecosystems, experiments impose perturbations and thus provide a more rapid and mechanistic understanding of an ecosystem's response to a perturbation (Hewitt *et al.* 2007). In addition, experiments are necessary to test predictions about nonlinear responses, non-additive interactions or tipping points that may arise from multiple concurrent perturbations (Scheffer & Carpenter 2003). There are many reasons that distributed experimental networks remain relatively rare, not the least of which may be the static (or declining) rates of funding for ecological research in many regions of the world. However, in light of world-wide alterations to ecosystems (Ellis *et al.* 2010), a globally coordinated experiment will produce insights unavailable using other approaches standard in our field. With careful planning, such experiments are positioned to address many other ecological questions, as well.

Here, we describe our experience developing and operating the Nutrient Network (NutNet, [nutnet.org](http://nutnet.org)), a globally distributed experimental network that arose from a grassroots, cooperative research effort (Box 1, also see Stokstad 2011). NutNet's origins began as an experimental solution to overcome the limitations of meta-analysis and synthesis. While meta-analyses can reveal the generality of patterns in existing data, they typically also highlight gaps in existing data that impede the resolution of unanswered questions. While

distributed experiments such as NutNet are not a panacea for understanding the factors affecting biological diversity and ecosystem structure and functioning, they complement our existing tools for understanding these processes and outcomes. For example, whereas a distributed experiment can provide insights into global generalities, many important ecological questions require more finely tuned, complex models and site-specific experiments (Table 1). Here, we clarify the design, goals and function of the Nutrient Network (a.k.a. NutNet) as a case study to provide guidance for the successful development of other such collaborative research networks. Whereas some lessons are highly general, others are not. However, we describe our experience with the goal of empowering others to employ the powerful, synergistic ecological tool of distributed experiments to examine global-scale ecological trends and responses.

## Designing a network

From the outset, we designed the Nutrient Network with clear scientific goals and questions, a simple, inexpensive and modular design, and room for site-level studies (Box 2). We agreed on protocols for core data collection to generate standardized, exactly replicated, directly comparable (e.g. sharing units, sample area and error structure) data from all sites. Because many ecological responses can emerge slowly, treatments and sampling are planned to continue for a minimum of 10 years. We started with a critical mass of sites; if no others joined our network effort, we were confident that our work alone (replicated at about six sites) would allow us to achieve novel scientific insights. With no central funding, participation was entirely voluntary; all site-level costs are covered by individual participating scientists. Finally, in recognition of the value of data (and costs incurred to each participating member) from each new location, we built our network with clear benefits for the scientists who contributed data to the effort.

### IDENTICAL TREATMENTS AND SAMPLING

The power of a distributed experiment lies in identical replication of treatments and sampling protocols at all sites; the lack of identical methods is often the greatest impediment to inference from meta-analysis (Table 1). From the outset of our planning, we agreed on experimental treatments, protocols and measurements that would be followed exactly at all sites

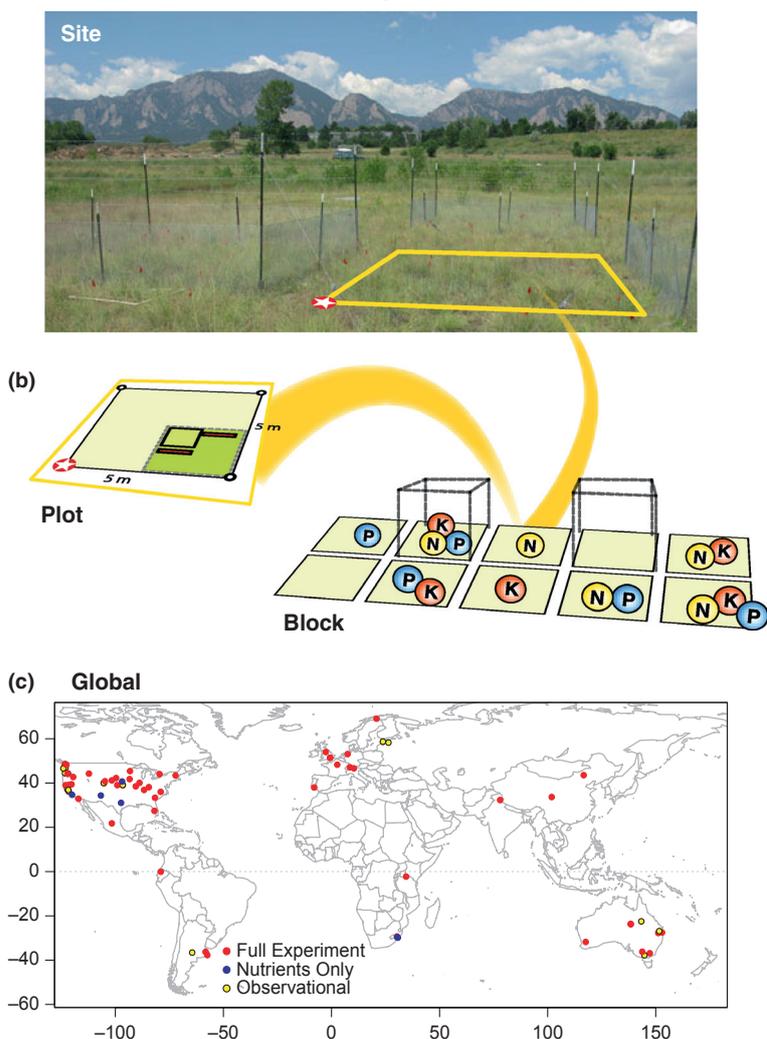
(see Appendix S1 for examples and explanations of exceptions). This was not always easy (we debated the details of one of our treatments at three different meetings across several

months), but the power of the network data set lies in the identical treatments, among-site replication, and standardized measurements.

### Box 1. NutNet experimental design: a case study

NutNet is composed of sites dominated by herbaceous vegetation in 17 countries on six continents. Sites represent the regional flora (e.g. tall-grass prairie, desert grassland, alpine meadow, agronomic pasture etc.) and are situated in a relatively homogeneous  $\sim 1000\text{-m}^2$  vegetation patch.

#### (a) Nutrient Network: a Globally-Distributed Experiment



(a) The NutNet study is a completely randomized block (environmental gradient) design with three blocks and 10 plots per block ( $N = 30$  total units/site). Each experimental unit is a  $5 \times 5$  m plot that is separated by at least 1-m walkways. Each plot is divided into four equal-sized subplots: one dedicated to core sampling (dark green, see *Core Sampling* below), one to additional site-specific or subnetwork studies and the last two for the future network-level research.

*Core Sampling.* In each plot, the core sampling  $2.5 \times 2.5$  m subplot is divided into four  $1 \times 1$  m permanent subplots, surrounded by a 0.25-m buffer. Within the core-sampling subplot, one  $1\text{-m}^2$  subplot is permanently marked for annual plant composition sampling; the other three are used for destructive biomass sampling. Core annual sampling includes clipping of total above-ground biomass of all plants rooted within two  $0.1\text{-m}^2$  strips ( $10 \times 100$  cm) for a total of  $0.2\text{ m}^2$ . These are sorted to live and dead (or further, e.g. forb, grass, moss at many sites), dried at  $60^\circ\text{C}$  to constant mass and weighed to the nearest 0.01 g. Leaves and current year's woody growth are collected from shrubs and subshrubs. Non-destructive biomass estimates based on plant allometry are used in a few highly fragile ecosystems. Light availability above and at ground level below the canopy is measured in the core subplot using a linear 1-m bar (e.g. Apogee Instruments, Inc., Logan, UT, USA). Areal cover is estimated to the nearest 1% for each species rooted in the core subplot; cover estimates include woody overstorey, litter, bare soil, rock and animal activity (e.g. digging). All core measurements are collected from all plots, annually at peak biomass.

Two 2.5 cm diameter by 10 cm depth soil cores, free of litter and vegetation, are collected from each plot prior to initiation of the experiment (Y0) and 3 years after treatment initiation (Y3). Soils from each plot are composited, homogenized, air-dried and shipped to a single laboratory

for analysis and long-term storage. Samples are assayed for % total C and % total N, extractable soil phosphorus, potassium and micronutrients, soil pH, soil organic matter and soil texture.

(b) The experimental treatments are applied at the scale of the  $5 \times 5$  m plots, as follows (Appendix S1 details examples of deviations):

*Fertilization treatments.* Three nutrient treatments (N, P and K plus micronutrients), each with two levels (control, added), are crossed in a factorial design, for a total of eight treatment combinations per block, to test multiple nutrient limitation on plant composition and ecosystem function (Table S1). Nutrient addition rates and sources are:  $10 \text{ g N m}^{-2} \text{ year}^{-1}$  as timed-release urea  $[(\text{NH}_2)_2\text{CO}]$  (see Appendix S1 for more N source details),  $10 \text{ g P m}^{-2} \text{ year}^{-1}$  as triple-super phosphate  $[\text{Ca}(\text{H}_2\text{PO}_4)_2]$ ,  $10 \text{ g K m}^{-2} \text{ year}^{-1}$  as potassium sulphate  $[\text{K}_2\text{SO}_4]$  and  $100 \text{ g m}^{-2}$  of a micronutrient mix of Fe (15%), S (14%), Mg (1.5%), Mn (2.5%), Cu (1%), Zn (1%), B (0.2%) and Mo (0.05%). N, P and K are applied annually; micronutrients were applied once at the start of the experiment to avoid toxicity.

*Fencing treatments.* A fencing treatment is crossed with the control and NPK treatments to assess the interactive effects of fertilization and food web manipulation on plant composition and ecosystem function. The 230-cm-tall fences restrict access by mid-to-large-sized above-ground mammalian herbivores ( $>50$  g). The lower 90 cm is surrounded by 1-cm woven wire mesh (hardware cloth) with a 30-cm outward-facing flange stapled to the ground to exclude digging animals (e.g. rabbits, voles), although not fully subterranean ones (e.g. gophers, moles). The upper fence is composed of four strands of tensioned wire strung at equal vertical intervals.

(c) The experimental design and sampling are replicated at grassland sites around the world (map of NutNet sites, August 2013). Most sites are contributing pre-treatment and all experimental data (full experiment); however, some sites have contributed only pre-treatment data (observational), and a few are applying only the nutrient addition treatments (nutrients only).

**Box 2. DIY – Key considerations for starting a do-it-yourself (DIY) experimental network**

- **Develop clear scientific goals and questions.** Articulating the scientific goals and core questions at the outset facilitates all subsequent decisions about network design.
- **Implement identical treatments and sampling.** Standardized, exactly replicated protocols for data collected from sites spread across regions and continents is one of the greatest contributions of such a network; deviations degrade the value of the data set. The protocols should be tested by your targeted audience and translated into the local language, where possible; culture, language and education can lead to a surprising diversity of interpretations of the same protocol.
- **Develop clear ground rules for participation.** Agreeing at the outset on key issues such as experimental design, how costs are paid, data sharing and authorship is important for maintaining a smooth long-term collaboration.
- **Use a simple, inexpensive design.** A low bar (cost and time) for basic participation means that more scientists will be willing and able to contribute at least these minimum data.
- **Use a modular design.** Some scientists will be interested in the contribution of some, but not all, data sets, so a modular design encourages data contribution from a wide variety of sites.
- **Use a flexible design with room for additional studies.** A flexible design with space for additional site- or regional-scale studies makes core participation more attractive to a broad group of scientists.
- **Start with a critical mass.** When designing a network, be sure that it will bear fruit even with data contribution solely from the core team.
- **Ensure clear benefits for participating scientists.** Clearly articulated scientific and professional benefits help busy scientists justify time for productive, new projects.
- **Plan for data management.** The scientific success of an experimental network depends on effective data management to facilitate analyses and novel insights (see Box 3).

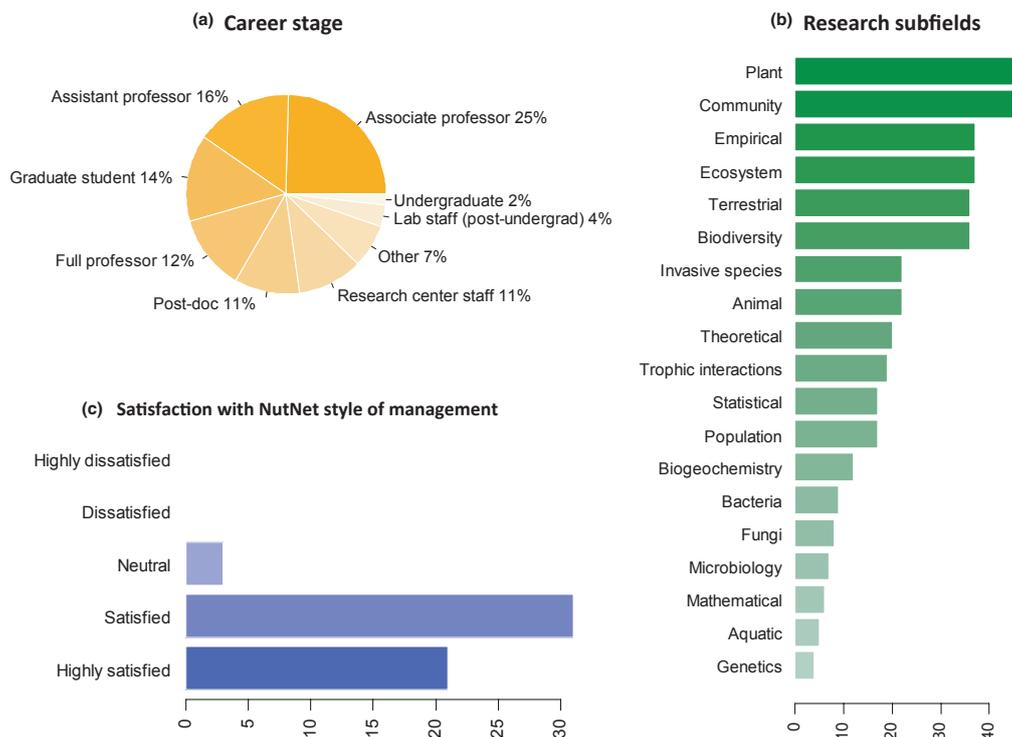
**CLEAR SCIENTIFIC GOALS AND QUESTIONS**

Clearly articulating scientific goals and core questions provides a reference point from which to make decisions about experimental design, core data and primary analyses. In our case, NutNet's origins were question-driven: we required standardized, exactly replicated, high-quality data collection to inform important ecological questions that we were not fully able to answer using traditional single-site experiments or meta-analysis. NutNet has two specific scientific goals. (i) To collect data from a broad range of sites in a consistent manner to allow effective quantitative comparisons of environment–productivity–diversity relationships among systems around the world and (ii) to implement a cross-site experiment requiring small investment of time and resources by each investigator, but quantifying community and ecosystem responses in a wide range of herbaceous-dominated ecosystems (i.e. desert grasslands to alpine meadows).

For the Nutrient Network, we laid out the following specific focal network research questions:

- 1 How general is our current understanding of the global patterns of grassland productivity and diversity?
- 2 To what extent are plant production and diversity limited by nutrients other than nitrogen or colimited by multiple nutrients in herbaceous-dominated communities?
- 3 Under what conditions do grazers or fertilization control plant biomass, diversity, composition and function?

These goals and theoretically grounded ecological questions are well served by a distributed experimental collaboration and allowed us to design treatments linking to both theory and current global environmental changes. We developed our experimental treatments to reflect current environmental change, hypothetical future states and ecological theory with the goal of using each of these elements to inform the others. By combining observations and experiments to address these questions at many sites across continents, we designed our experiments and sampling to inform a general, mechanistic understanding of our experimental factors and interactions while also quantifying the role of natural environmental



**Fig. 1.** A July 2013 survey of Nutrient Network scientists (54 respondents) demonstrates diversity in (a) career stages, (b) research subfields and (c) overall satisfaction with the NutNet management structure and style described in this paper.

variability in space and time in determining these functional relationships (Hewitt *et al.* 2007).

#### GROUND RULES FOR PARTICIPATION

‘Thou shalt play well with others’ is the primary ground rule in the Nutrient Network and one that has selected for a highly cooperative network of collaborators. To meet our minimum bar for network participation, collaborators agree to exactly follow the design, core treatments and core measurements determined by the discussions among the core project partners, and collaborators must provide data to the central data base for other network members to use. Our ground rules also define the process of starting new research projects using the NutNet infrastructure, using data and authoring papers.

#### SIMPLE, INEXPENSIVE DESIGN

We sought to design an experiment that would represent minimal expense (e.g. time, equipment or sample processing) for any single site, but would create a novel, valuable data set for answering major questions for the field. With dedicated funding, we would likely have designed and implemented our experiment differently (e.g. targeting specific locations and using more sophisticated treatments). In the absence of dedicated central funding, achieving our scientific goals required creativity and flexibility in defining our project scope. For example, we spent some time considering whether a rainout shelter or fencing treatment would create a sufficiently novel

and important network-scale data set that would justify the substantial additional expense. We opted to include a fencing treatment (see Box 1) and designed this to balance time, expense, effectiveness and maintenance considerations. With careful attention to these constraints, our NutNet design required low annual financial and labour investments.

Our experimental design included minimal within-site replication (three replicates per treatment). The low within-site replication minimizes costs and sampling effort for any single investigator; the statistical power of this network design resides in its among-site replication.

#### MODULAR DESIGN

Each additional data point increases the value of a distributed experiment. To maximize the number of sites participating in data collection, we planned for three levels of participation, in the form of network-scale data sets. The contribution requiring the least effort was a one-time sampling (30 plots of plant composition, plant biomass, soil collections and light interception) that required 2–3 days of time by three people knowledgeable about the local flora. These observational data have proven extremely valuable, providing new insights into invasive species (Firn *et al.* 2011), litter decomposition rates (O’Halloran *et al.* 2013) and diversity and productivity relationships (Adler *et al.* 2011; Grace *et al.* 2012).

Contribution beyond this entailed annual sampling and application of a factorial combination of three nutrient addition treatments (‘nutrient experiment’). This treatment cost is approximately US\$300 per year per site, an achievable price

for scientists at any career stage and for scientists in most countries of the world. We designed our final data set to require the addition of only two more plots in each block with fences around them ('fence experiment'), which added a one-time supplies expense of approximately US\$3000 per site, also achievable by most scientists. Individual investigators have paid for costs of experimental set-up and maintenance, including travel and labour, typically from start-up funds or local site small, research grants (see below). Given our 10-year plan for experimental treatments and sampling, the control plots alone will produce a valuable long-term data set (Hewitt *et al.* 2007). Although most sites contribute to all three data sets, some contribute to two, and we receive a small but steady stream of observational data from sites around the world (Box 1).

#### FLEXIBLE DESIGN

Understanding and predicting the effects of global anthropogenic changes requires interdisciplinary teams (Metzger, Lee-mans & Schroter 2005). To this end, we sought to attract scientists with a variety of academic interests by designing flexibility into the experiment. While we required a core set of experiments and measurements, we also designed the plots to allow individual investigators to answer site-specific questions, depending upon investigator interests, time and resources (Box 1). Some investigators implemented additional blocks or treatments (e.g. Ziter & MacDougall 2013). Many sites participated in additional multi-site sampling efforts (see Add-on Studies). Our approach has been quite effective in attracting scientists from a diversity of career stages and subfields (Fig. 1).

#### CLEAR BENEFITS FOR PARTICIPATING SCIENTISTS

The value of a distributed experimental data set increases with participation, which, in turn, increases when participants directly benefit from their contributions of time and money. We designed our network with a variety of clear benefits. First, we implemented a 'pay to play' system of data contribution: data sharing provides benefits of data access and opportunities for co-authorship on network publications. For the first network publication using each data set, we deemed that data contribution was sufficient for co-authorship. For all subsequent papers, solely providing data was insufficient for authorship; data contribution in addition to substantive contribution to idea development, analysis or writing is sufficient for co-authorship on network publications. This is tracked through an authorship form required for all publications (see Appendix S2 for authorship rubric).

Along with benefits of co-authorship, contribution of data provides early data access; even after a single year of data collection, the network data represented a powerful data base for participants to use for a broad variety of analyses. Finally, participation in the experiment, in particular, provides a collaborative team and experimental infrastructure for additional cross-site research (see Add-on Studies) and creates a valuable

networking opportunity for junior scientists in the network (Fig. 1). As an added incentive for participants, our design included subplots for site- or regional-scale studies, with the only constraint being that any additional treatments could not threaten the integrity of the long-term core data set (e.g. seed addition, which could confound future species diversity, population genetic or invasion questions). To ensure concurrent benefits for the broader scientific community, data associated with network studies are made publicly available at the time of publication, and we remain open to requests for data and collaboration by those outside the network.

#### ADD-ON STUDIES

At the outset of this endeavour, we envisioned these plots as an experimental platform, interesting for many ecological questions beyond our core foci. To this end, we encouraged network members to design additional studies that were easy for any site to implement and placed the burden of cost, time and effort on the laboratory initiating the study. For example, the earliest add-on study involved mailing ion resin bag 'kits' to participating sites to monitor soil nutrient fluxes; researchers at receiving sites deployed the resin bags and mailed them back to the initiating investigator's laboratory for analysis. The bar for participation in this type of add-on study is relatively low, and the benefit of participation, co-authorship, is significant; benefits to initiating investigators are also substantial. Add-on studies have been extremely successful for our network, leading to additional measurements and experiments replicated at up to 40 sites.

#### Network management

Whereas much of the scientific effort of the network can be distributed, network data management must be centralized to most effectively meet the ongoing scientific goals of the group. While the growth of our network has increased our power to address a variety of questions, both the ongoing network growth and the increasing diversity of data have required dedication of a significant portion of a position to data (and people) management since 2007. This raises a few important considerations that should be considered at the outset.

#### MANAGEMENT STYLE

A network's management must be structured to support the network's mission; however, there exist a wide variety of management styles that may be appropriate for different personalities, organizations and goals. Based on the initial goal of NutNet requiring a minimal investment of time from all participants, our lack of centralized funding and the personalities of the scientists initiating our network, we sought to be highly flexible, self-organized, low budget and egalitarian while retaining a focus on the scientific products of the data. Minimalist management of our network has worked well, to date. Our grassroots research effort is coordinated and maintained by a few central group participants. In particular, day-to-day

management of the network is led by a network manager, a postdoctoral scientist chosen from a national search. The network manager has primary responsibility for data management and serves as a point person for inquiries about site set-up and maintenance and recruitment of new members and sites. The manager works closely with two network coordinators (among the group that envisioned and designed this network; this paper's authors) who communicate with network scientists, organize and run network meetings, enforce network standards for experiment implementation, resolve disputes and communicate about the network with the scientific community. Major decisions about the network, such as the long-term scientific direction of network research, is determined by a steering committee (currently ~12 scientists from four countries), including most of the original group that envisioned and designed this network and other self-selected network participants who have chosen to contribute significant time to network-wide NutNet management issues. We have few committees; the most active is the Authorship Committee that tracks the data sets and analyses being used by network members and works to facilitate communication between groups interested in similar analyses. Our consensus-based organization has experienced remarkably few conflicts, to date, but these have been resolved through mediation by the authorship or steering committee, depending on the nature of the conflict. While the structure and management is one among many options for other networks, NutNet participants are satisfied with this approach for our network (Fig. 1c).

#### FUNDING

Although the science of the network is funded at the site level by individual investigators, the network coordinators have sought funding for the network manager position as well as funds for meetings. These two components, focused on data integrity, access and analyses, are keys to achieving the scientific goals of the network. In general, we have found significantly more opportunities for funding of meetings; funding opportunities for the network manager position are far rarer. Our funding for this position has come through the National Science Foundation Research Coordination Networks (4 years), the University of Minnesota Institute on the Environment (3-5 years) and the National Science Foundation Long Term Ecological Research funding for the Cedar Creek LTER site (2-5 years). Networks with other research foci will likely find other funding sources for data management.

#### NETWORK MEETINGS

The primary role of NutNet collaborators is to contribute to the network's distributed research effort. Advancement of the field requires communication among network scientists both electronically and in person. To this end, we hold annual working meetings to bring together subsets of the NutNet community to participate in discussions about the most important issues to address using these data, plan and implement analy-

ses, write manuscripts and develop future projects. Students and early career scientists participate in all aspects of network research, including setting up and sampling sites, leading add-on studies and, importantly, contributing to and leading network analyses and manuscripts.

In-person interactions are a critical collaborative tool, but they represent a significant expense – in the case of NutNet, these costs are often substantially greater than the costs associated with implementing the experiment. We use video conferencing, but given the global distribution of scientists, finding a time when all callers are able to effectively participate is a major challenge. Video conferencing also can exclude people where technology is poor. In our experience, however, meeting expenses can be reduced by hosting meetings in locations with reduced lodging costs (e.g. university dormitories, field stations), and a variety of sources exist for funding meetings.

#### AUTHORSHIP

Because academic careers are built, in part, on authorship and intellectual contributions, authorship considerations are a critical, ongoing component of network management. We seek to appropriately acknowledge the efforts of each scientist involved in this work. Although long author lists are common in other subfields of biology (e.g. Human Genome Project), they are less common in ecology. We have adopted a policy that requires every paper submitted using NutNet data to include an author contribution table to clearly identify author contributions in a transparent and consistent manner (Appendix S2). This has the dual effect of making ourselves internally accountable while also justifying each person's role as an author to the larger scientific community.

#### LINKING WITH LONG-TERM DATA SETS

There is a long history of work in grasslands, and we wanted to link our large-scale network data to long-term data sets (collected using a variety of protocols) in this ecosystem. To this end, we identified sites with a long history of conceptual contributions in ecology (e.g. Cedar Creek LTER, Konza LTER, Silwood Park, Serengeti National Park etc.) and solicited participation in our network from researchers at these sites. New networks will likely find such synergy with existing efforts.

#### NETWORK DATA MANAGEMENT

The strength of distributed experiments lies in analysing data generated by standardized, exactly replicated protocols, which in turn requires merging data from diverse sites into uniform data sets. The importance of site geographic location and the hierarchical experimental design of our network led us to implement a relational data base model (Box 3). Our model associates observed data with the spatial location in which it was observed (typically, a subplot). The spatial location is then nested within broader spatial entities (plots, blocks and sites), which can have their own associated information (e.g. treatment, establishment year, and latitude and longitude); such

**Box 3.** Data management for a distributed experimental network

- **Relational data bases are a good fit for network data.**

With a relational approach, data are only stored and edited in a single place in the data base. Thus, only one row in a data table within the data base stores the plot number and treatment for a given plot, or subplot-scale response data. These single pieces of information are then linked ('related') to other pieces of data through identifiers, as one-to-one (a live biomass value for a year in a given subplot), many-to-one (four subplots in a plot, 30 plots in a site) or many-to-many relationships. This structure elegantly handles the multi-scale, interrelated nature of distributed experiments, reduces storage costs and transcription error, and increases standardization and flexibility of data integration (Borer *et al.* 2009).

- **Relational data bases are inexpensive and need not be complicated.** Many software options (including the nearly ubiquitous desktop software Microsoft Access) facilitate construction of data models using menu-driven, visual tools. Software tools such as MySQL and PostgreSQL (SQL stands for Structured Query Language, the interface language of many data bases) are available free of charge for download and installation, and many universities offer low-cost data base hosting and consulting.

- **Data integration can be automated but requires hands-on attention.** Standard data submission templates, parsing scripts and internet-based data submission can reduce errors from collection to submission and integration. Still, the nature of ecological data, including taxonomic identities and observations with temporally varying parameters, requires human oversight to ensure a high-quality data set. For a distributed experiment, quality is most efficiently achieved by a data manager with the knowledge of data structures and standardization. Centralized data reconciliation also makes network data contribution more feasible for those lacking skills or patience for data integration and QC.

- **Metadata is crucial to understanding data inside and outside the network.** Associating metadata, or data about data, with experiments has been more commonly advocated than implemented in ecology (Michener *et al.* 1997). Even within a single laboratory, metadata provides information about each column of data: 'What are the units?' 'What are the derivations?' 'Why are data missing?' These and many other specifications are critically important for the synthesis of many site-level observations and creation of a single data set. Idiosyncrasies that are apparent or even obvious within a given site disappear the moment that data from that site are nested in a network data set. Assigning and maintaining metadata permanently associates these important details with the data (Borer *et al.* 2009).

- **Authorship and project proposals can be managed using data bases.** Proposed and active research projects also can be managed in a relational data base. In NutNet, proposals to create an analysis or manuscript using network data are submitted through an online form, which is automatically forwarded to the authorship committee. This includes information about the data subset and planned analyses. Tracking this in a data base format can be helpful because network members can propose multiple papers; papers generally have many network authors, and projects use various overlapping subsets of data. New network collaborators can search this data base, contact lead authors, contribute to existing analytical efforts or propose new ones.

information can also be geo-referenced. A relational approach facilitates data extraction for a variety of questions, even unanticipated ones. Diverse information can be related via common spatial association: plant species observed in a plot can be linked to soil nutrient content and light availability, although these different data types are stored in separate tables. Add-on projects in a new table can be easily linked to core data via a plot or subplot identifier. Finally, spatial identifiers, especially site geographic coordinates, allow the association of fine-scale data with the vast diversity of publically available ancillary data, such as climatic variables, soil classifications and remote sensing information. Such ancillary data often come from a single source or model, providing a fully standardized global data set for all experimental subplots.

Two major challenges have accompanied the integration of site-level NutNet data into a single data base. First, the idiosyncratic variation in data entry or data management practice at a single site can make data integration difficult, or worse, can introduce subtle unit errors (e.g. reporting plant biomass in the collection units of  $\text{g } 0.2 \text{ m}^{-2}$  versus the transformed units  $\text{g m}^{-2}$ ). We address this with standardized electronic data submission forms and with quality control (QC) measures such as outlier analysis. Not every network member has the time, resources or inclination to follow precise data submission instructions, resulting in significant labour investment in data integration. Secondly, as NutNet involves multiple aspects of plant species diversity, taxonomic resolution of plant names is necessary to ensure 'apples-to-apples' comparisons across sites and regions. We do not seek to resolve controversy over botanical nomenclature, but rather to assign each plant taxon in our data base to a definite entity, which requires having a single, standard source for botanical names. While recently developed online tools (e.g. the Taxonomic Name Resolution Service, <http://tnrs.iplant-collaborative.org>) have facilitated the automation of this task for many data sets, these services work best for the well-studied North American and Neotropical flora. Because of NutNet's global reach, we chose to use a global reference checklist compiled by [www.theplantlist.org](http://www.theplantlist.org), a collaboration among Royal Botanic Gardens at Kew, Missouri Botanical Garden and various other name resolution efforts for specific higher taxa. Using this resource, we can match plant species names and assign all observations to a unique global taxon, while keeping local site names stored in a separate table.

## Conclusions

Significant advances in our ability to predict the response of ecological systems to perturbations require an understanding of the role of environmental context and multi-factorial relationships, which in turn requires standardized observations and experimental manipulations at many sites. Thus, global science will be an essential tool for grappling with many of the most important ecological questions. Globally distributed experimental science represents an emerging and powerful approach that promises more effective development of a predictive understanding of the impacts of global change on eco-

systems. The benefits of a globally distributed experimental network are that it does not need to be prohibitively expensive or time-consuming and is not limited to senior scientists or those in countries where science is well funded. In addition, because the approach is grassroots in nature, critical data can be collected relatively rapidly and the experimental platform can be used to respond to new and emerging issues. Development of and participation in such a research collaboration can be highly productive and can lead to scientific synergy and high-impact advances for the field that would otherwise be unachievable. However, even in the efficient, relatively low-cost model we present here, data management and network coordination must be centralized to ensure data integrity and scientific impact of the network. Meetings, for brainstorming, analyses and writing, are also critical for the advancement of science using network data.

The type of distributed experimental framework we describe here provides another powerful tool for overcoming some of the obstacles we face in seeking generality and addressing the grand challenges for the environmental sciences identified by the National Academy of Sciences (NAS 2001). Our example, the Nutrient Network, seeks to scientifically address some of the elements of a single grand challenge for our field – understanding the generality of processes determining biological diversity and ecosystem structure and functioning and quantifying the conditions under which these relationships are strongest – using this approach. While this approach is not tailored to maximize site-level insights, it is an excellent approach for generating a general, predictive understanding of the functional relationships controlling ecological responses. The National Academy of Sciences laid out several other grand challenges for the environmental sciences, including furthering our understanding of natural and anthropogenic drivers of Earth's biogeochemical cycles, climate variability, hydrologic forecasting, infectious disease and the environment, and land-use dynamics (NAS 2001). Seeking generality and quantifying local and regional contingencies in each of these research areas – and many others – could be effectively approached with such a distributed experimental framework.

## Acknowledgements

The authors wish to thank all of the individuals involved in the Nutrient Network distributed experiment for their ongoing contributions to developing this approach as an effective scientific tool. Nutrient Network is funded at the site scale by individual researchers. Coordination and data management have been supported by funding to E. Borer and E. Seabloom from the National Science Foundation Research Coordination Network (NSF-DEB-1042132) and Long Term Ecological Research (NSF-DEB-1234162 to Cedar Creek LTER) programmes and the Institute on the Environment (DG-0001-13).

## References

- Adler, P.B., Seabloom, E.W., Borer, E.T., Hillebrand, H., Hautier, Y., Hector, A. *et al.* (2011) Productivity is a poor predictor of plant species richness. *Science*, **333**, 1750–1753.
- Borer, E.T., Seabloom, E.W., Jones, M.B. & Schildhauer, M. (2009) Some simple guidelines for effective data management. *Bulletin of the Ecological Society of America*, **90**, 205–214.
- Denny, M. & Benedetti-Cecchi, L. (2012) Scaling up in ecology: mechanistic approaches. *Annual Review of Ecology, Evolution, and Systematics*, **43**, 1–22.

- Ellis, E.C., Goldewijk, K.K., Siebert, S., Lightman, D. & Ramankutty, N. (2010) Anthropogenic transformation of the biomes, 1700 to 2000. *Global Ecology and Biogeography*, **19**, 589–606.
- Elmendorf, S.C., Henry, G.H.R., Hollister, R.D., Bjork, R.G., Bjorkman, A.D., Callaghan, T.V. *et al.* (2012) Global assessment of experimental climate warming on tundra vegetation: heterogeneity over space and time. *Ecology Letters*, **15**, 164–175.
- Firn, J., Moore, J.L., MacDougall, A.S., Borer, E.T., Seabloom, E.W., Hillebrand, J. *et al.* (2011) Abundance of introduced species at home predicts abundance away in herbaceous communities. *Ecology Letters*, **14**, 274–281.
- Fraser, L.H., Henry, H.A.L., Carlyle, C.N., White, S.R., Beierkuhnlein, C., Cahill, J.F. *et al.* (2012) Coordinated distributed experiments: an emerging tool for testing global hypotheses in ecology and environmental science. *Frontiers in Ecology and the Environment*, **11**, 147–155.
- Grace, J.B., Adler, P.B., Seabloom, E.W., Borer, E.T., Hillebrand, H., Hautier, Y. *et al.* (2012) Response to comments on "Productivity is a poor predictor of plant species richness". *Science*, **335**, 1441.
- Gurevitch, J. & Mengersen, K. (2010) A statistical view of synthesizing patterns of species richness along productivity gradients: devils, forests, and trees. *Ecology*, **91**, 2553–2560.
- Harmon, M.E., Silver, W.L., Fasth, B., Chen, H., Burke, I.C., Parton, W.J., Hart, S.C., Currie, W.S. & LIDET (2009) Long-term patterns of mass loss during the decomposition of leaf and fine root litter: an intersite comparison. *Global Change Biology*, **15**, 1320–1338.
- Hector, A., Schmid, B., Beierkuhnlein, C., Caldeira, M.C., Diemer, M., Dimitrakopoulos, P.G. *et al.* (1999) Plant diversity and productivity experiments in European grasslands. *Science*, **286**, 1123–1127.
- Hewitt, J.E., Thrush, S.F., Dayton, P.K. & Bonsdorff, E. (2007) The effect of spatial and temporal heterogeneity on the design and analysis of empirical studies of scale-dependent systems. *American Naturalist*, **169**, 398–408.
- Levin, S.A. (1992) The problem of pattern and scale in ecology. *Ecology*, **73**, 1943–1967.
- Metzger, M.J., Leemans, R. & Schroter, D. (2005) A multidisciplinary multi-scale framework for assessing vulnerabilities to global change. *International Journal of Applied Earth Observation and Geoinformation*, **7**, 253–267.
- Michener, W.K., Brunt, J.W., Helly, J.J., Kirchner, R.B. & Stafford, S.G. (1997) Nongeospatial metadata for the ecological sciences. *Ecological Applications*, **7**, 330–342.
- Mulholland, P.J., Tank, J.L., Webster, J.R., Bowden, W.B., Dodds, W.K., Gregory, S.V. *et al.* (2002) Can uptake length in streams be determined by nutrient addition experiments? Results from an interbiome comparison study. *Journal of the North American Benthological Society*, **21**, 544–560.
- National Academy of Sciences, Committee on Grand Challenges in Environmental Sciences. (2001) *Grand Challenges in Environmental Sciences*. The National Academies Press, Washington, DC.
- O'Halloran, L.R., Borer, E.T., Seabloom, E.W., MacDougall, A.S., Cleland, E.E., McCulley, R.L. *et al.* (2013) Regional contingencies in the relationship between aboveground biomass and litter in the world's grasslands. *PLoS ONE*, **8**, e54988.
- Scheffer, M. & Carpenter, S.R. (2003) Catastrophic regime shifts in ecosystems: linking theory to observation. *Trends in Ecology & Evolution*, **18**, 648–656.
- Stokstad, E. (2011) Open-source ecology takes root across the world. *Science*, **334**, 308–309.
- Ziter, C. & MacDougall, A.S. (2013) Nutrients and defoliation increase soil carbon inputs in grassland. *Ecology*, **94**, 106–116.

Received 5 June 2013; accepted 30 September 2013

Handling Editor: Robert Freckleton

## Supporting Information

Additional Supporting Information may be found in the online version of this article.

**Appendix S1.** Distributed experiments require flexibility: examples of deviations from the original NutNet protocols

**Appendix S2.** NutNet authorship guidelines and rubric, working document (July 2013).

**Table S1.** Nutrient addition treatments and rates of application.