CellPress
Trends in
Ecology & Evolution

Opinion

Uniting Experiments and Big Data to advance ecology and conservation

Robert McCleery [1,6,*] Robert Guralnick, [2,6] Meghan Beatty, [1] Michael Belitz, [2] Caitlin J. Campbell, [3] Jacob Idec, [2] Maggie Jones, [4] Yiyang Kang, [5] Alex Potash, [1] and Robert J. Fletcher Jr [1,6]

Many ecologists increasingly advocate for research frameworks centered on the use of 'big data' to address anthropogenic impacts on ecosystems. Yet, experiments are often considered essential for identifying mechanisms and informing conservation interventions. We highlight the complementarity of these research frameworks and expose largely untapped opportunities for combining them to speed advancements in ecology and conservation. With nascent but increasing application of model integration, we argue that there is an urgent need to unite experimental and big data frameworks throughout the scientific process. Such an integrated framework offers potential for capitalizing on the benefits of both frameworks to gain rapid and reliable answers to ecological challenges.

Highlights

The use of 'big data' is increasingly championed as a scientific framework to help solve our global conservation crisis. Yet experiments are also advocated as essential for identifying mechanisms and interventions needed to solve this crisis.

Big data can provide the scaffolding for documenting patterns across scales, whereas experiments can inform novel situations and provide the means for conservation actions.

The complementarity of these frameworks provides untapped opportunities for leveraging the strengths of both frameworks through an integrative process.

We outline an integrated framework that can address a range of conservation problems, including forecasting, ecological novelty, scaling, and generalization of conservation actions.

A scientific revolution in need of balance

New information and perspectives are revolutionizing the way scientists address ecological and conservation problems. With the growing availability of earth monitoring tools, digital sensors and community science, coupled with flexible statistical approaches, **big data** (see Glossary) are increasingly championed to help solve global conservation problems and answer long-standing questions in ecology [1,2]. Harnessing a growing number of large and diverse data streams of ecological **observations**, or big data, ecologists have been able to explore and reveal ecological patterns (e.g., climate change effects, species declines) on unprecedented temporal and spatial scales [1]. While this **Big Data Framework** (Box 1) shares some similarities with attributes of **observational studies**, the scope of large data streams and their interpretation carry new challenges [3]. In particular, there are an increasing number of examples where the conclusions from big data can be misleading [3]. This has led to concerns that conclusions from the Big Data Framework are: (i) not accurate at actionable scales [4]; (ii) not able to elucidate causal mechanisms [4,5]; and (iii) limited in their ability to predict future ecological states [6].

Meanwhile, some ecologists continue to advocate for the **Experimental Framework** (Box 1) to produce an understanding of ecological processes and their relevance to conservation [7,8]. The use of well-planned **experiments** has produced ecology's core understandings of population limitations, species interactions, the role of biodiversity in ecosystem functioning and has identified specific causes and solutions for environmental degradation [9–11]. Yet, experiments are often limited to specific taxonomic groups and by their spatial and temporal scope. A recent surge in the coordination of field experiments provides one means to address these issues [12,13], but no matter how broadly deployed, concerns remain that the implementation of experiments can be slow, inefficient, costly, and unable to capture the breadth and complexity of our conservation crisis [13–15].

Some scientists still advocate for their preferred frameworks [13–16]. For example, while arguing for the use of predictive models with observational data, Currie [13] notes that, 'Mechanisms that

[1]Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, FL 32618, USA
[2]Florida Museum of Natural History, University of Florida, Gainesville, FL 32618, USA
[3]Department of Biology, University of Florida, Gainesville, FL 32618, USA
[4]School of Natural Resources and the Environment, University of Florida, Gainesville, FL 32618, USA
[5]Department of Environmental Engineering Sciences, University of Florida, Gainesville, FL 32618, USA
[6]These authors contributed equally to this work

*Correspondence: ramccleery@ufl.edu (R. McCleery).

 https://doi.org/10.1016/j.tree.2023.05.010

Box 1. Scientific frameworks in ecology and conservation

Ecology and conservation apply a variety of epistemologies to address problems. Big data has grown from the availability of large data streams and analytical tools. We define the Big Data Framework as the generation, design, analysis, and interpretation of big data as an epistemology for science [48] (Figure IA). Studies that use big data share similarities with observational studies, but differ in design and execution, where big data studies typically do not have design-based inference because they often use nonprobability samples [3]. By contrast, experiments are motivated by a desire to establish causality and identify the processes driving patterns. The Experimental Framework typically uses a hypothetico-deductive approach, using manipulations to identify mechanism or address perturbations used for conservation action (Figure IA).

The benefits of each framework are clear (Figure IB). The Big Data Framework can harness large amounts of data at low cost to the end-user, potentially increasing the precision of estimated patterns. Furthermore, it has tracked changes in biodiversity across unprecedented scales, showing losses in the diversity of animals [49–51] and identifying areas of degradation for restoration [52,53]. It has also illuminated climate-induced changes in phenology [54], species distributions [55], and disturbance regimes [56]. By contrast, the Experimental Framework has been used to evaluate interventions such as fire management [57], payment of ecosystem services [58], and removal of invasive species [59]. It has also determined the mechanisms (e.g., light pollution [11] and introduced predators [60]) generating species declines.

In isolation, each framework has limitations due to different types of uncertainty. Inferential uncertainty revolves around the problem of reliable knowledge when faced with bias, correlation, and external validity. The Big Data Framework often uses unstructured and unplanned data, potentially leading to bias and loose linkages to questions [61]. Correlations and induction are useful for making associations and generating hypotheses but can be limited in examining complex causal questions. The external validity of results from the Experimental Framework may be limited because outcomes may not be relevant to larger populations or spatial extents. Individually, the Experimental and Big Data Frameworks are both affected by extrapolation uncertainty that comes from applying conclusions across space and time. The Experimental Framework is typically implemented at fine grains and small extents, limiting its application to large-scale problems. By contrast, the Big Data Framework is frequently implemented at coarse grains and large extents, yet downscaling these approaches to finer scales, where conservation actions occur, is often challenging.

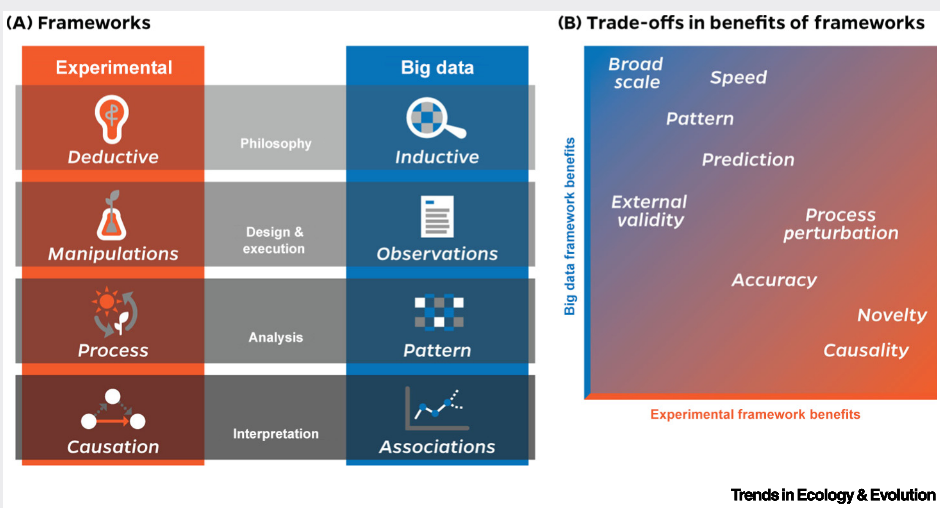


Figure I. Contrasting the attributes of the Experimental and Big Data Frameworks and their complementarity. (A) Experimental and Big Data Frameworks often differ in philosophy, design and execution (or implementation), analysis, and interpretation. For instance, experimental frameworks use manipulations that generate probability samples, whereas big data frameworks typically use high volume, nonprobability samples from observations. (B) Based on these differences, the benefits of each framework differ, such that there are trade-offs when using each framework in isolation. Yet, the shortcomings of each can be mitigated by leveraging and integrating these frameworks throughout the scientific process.

are statistically detectable in experimental systems may contribute very little to the variation of nature'. Alternatively, Wiersma [8] counters that, 'Experiments are at the heart of what it means to "do science"'. Still others have a keen interest in integrating Big Data and Experimental Frameworks to overcome their shortcomings [1,17] and better address environmental challenges.

Glossary

Big data: data with typical characteristics such as high volume, collected at high frequency and contains a complex mix of data types. Often includes nonprobability samples.
Big Data Framework: the generation, design, analysis, and interpretation of big data as an epistemology for science.
Causal modeling: models that are constructed for gaining causal inference, by focusing on potential cause-and-effect relationships from observation (e.g., using directed acyclic graphs).
Counterfactual modeling: models that are constructed to predict how a causal factor is a necessary factor without which the outcome (e.g., treatment success) would not have occurred.
Data integration: making data from different sources usable and useful as a cohesive product by increasing interoperability.
Design integration: linking different scientific methods, such as experimental treatments with nonprobabilistic sampling, in study designs.
Experiment: a deliberate manipulation of one or more variables done in such a way that other variables are held constant or rendered unimportant through suitable study design. The three cornerstones of experiments are randomization, replication, and manipulation.
Experimental Framework: the use of manipulations to identify mechanism or address perturbations used for conservation action typically employing a hypothetico-deductive approach.
Hybrid dataset: a dataset that combines observational with experimental data.
Model integration: using models that combine different sources of data under a common framework or the creation of 'meta-models' that combine component models.
Nonprobability samples: data that are sampled whose sampling mechanisms are unknown. Common in opportunistic data used in the Big Data Framework.
Observational study: a study that lacks at least one of the three cornerstones of experimentation, particularly manipulation, but study design is structured such that it includes probability samples for a treatment or factor(s) of interest.
Observation: any sort of measurement of ecological phenomena.

There are well-known limitations of both the Experimental and Big Data Frameworks (Box 1), yet many of these could be overcome by leveraging each of their strengths. The development of modeling approaches that can integrate big, experimental, and other types of observational data has been a critical first step [1,17–19]. However, **model integration** alone is not sufficient to address many challenges. We argue that combining Experimental and Big Data Frameworks throughout the scientific process (i.e., from conceptualization to interpretation and dissemination) will best address our conservation crisis on the scales where ecological processes can be altered and managed. Here, we provide a vision for uniting big data and experimentation to more rapidly and reliably advance ecology and conservation.

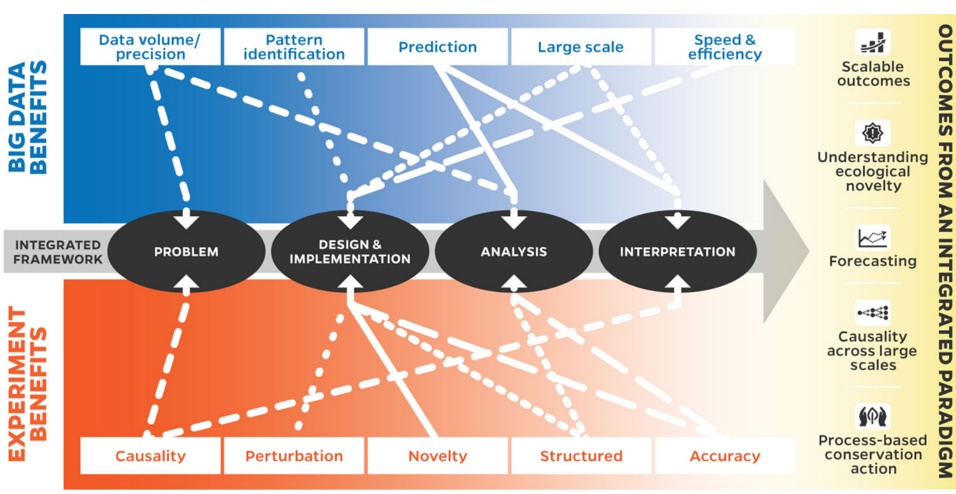
Probability samples: data that are sampled whose sampling mechanisms are known or designed, such as from structured data. Common in observational and experimental studies.

Complementarity in frameworks generate opportunities

Conservation strategies require synthesizing a complexity of information for effective outcomes. The Big Data Framework can often provide a foundation for documenting and monitoring patterns of biodiversity across spatial scales (from local to global [1]). The Experimental Framework can deliver direct assessments of perturbations that are relevant for conservation interventions (e.g., disturbance, restoration) and can inform the understanding of novel situations. For example, recent syntheses using a Big Data Framework highlight that climate change has not been the primary driver of extinction risk [20]. Yet, the counter argument is that patterns uncovered from the recent past may not inform expectations for a novel climate future [21], an issue that can be addressed with the Experimental Framework. Both arguments are correct, and yet neither can be fully addressed in the absence of each other. Combining these frameworks could provide value by balancing the limitations of one with the strengths of the other.

An Integrated Framework for the future of ecology and conservation

To fully realize the benefits of Experimental and Big Data Frameworks, integration is needed throughout the scientific process. Although the process and workflows of these frameworks are not identical, they share components that provide a means for integration. Both frameworks



Trends in Ecology & Evolution

Figure 1. A workflow of an Integrated Framework unifying Big Data and Experimental Frameworks. The benefits (white boxes) of Big Data (blue) and Experimental (orange) Frameworks can be leveraged to inform each stage (black circle) of the scientific process in an Integrated Framework. White lines extend from white boxes in different patterns unique to each benefit. Lines represent the most likely stage of the scientific process where these benefits can be harnessed. Utilizing the benefits of each framework throughout the scientific process will lead to a host of improved outcomes (in yellow) for ecology and conservation. See worked examples in Boxes 2 and 3.

include hypothesis or question generation, design and implementation, analysis, and interpretation [22,23]. Integration can occur for each component of these frameworks (Figure 1), but in practice it will be necessary to weigh the benefits of each framework over the other at different stages of the scientific process. We discuss integration for each component; see Boxes 2 and 3 for worked examples.

Question and hypothesis generation

Hypothesis and question generation frequently proceed by probing observed patterns [24]. While experiments have often emphasized hypothetico-deductive philosophies and the Big Data Framework has emphasized more inductive logic, there are opportunities to have these approaches feed off each other. The Big Data Framework can be used to speed the discovery of patterns to refine and expand hypotheses for testing by experiments [25]. For instance, experiments on predation risk have provided invaluable insights on the roles of consumptive versus nonconsumptive effects [26], but how such interactions play out among individuals in space and time remains challenging to understand. This leads to key questions regarding behavioral contact processes that big data from per-second readings of high-resolution GPS tags can isolate in ways not possible from other types of information [15]. High-resolution data can also reduce biases that may arise when interpreting behavioral interactions with coarse-scale observations, such as the underestimation of anti-predator responses by elk (*Cervus canadensis*) to wolves (*Canis lupus*) [27]. Coupling high-resolution movement data with fear experiments [28] may shed new insights into landscapes of fear and their consequences.

Design and implementation

The design and implementation of science can also be improved through **design integration**. The Experimental Framework can harness gradients identified from the Big Data Framework

Box 2. Example of an Integrated Framework to inform phenological changes from environmental changes

Problem

Forecasting plant and animal phenology over short and long timescales remains challenging, despite its strong sensitivity to environmental changes, such as warming [62]. It is poorly understood how much short-term climate variability drives the timing of subsequent phenological events (Figure I). Although previous work on plant phenology based on satellite remote sensing suggested marked impacts [63], initial big data analyses utilizing coarse spatial grain historical data for North American lepidopterans (butterflies and moths) and temporally fine-grain (daily) historical climate data have shown strong associations between unusually warm or cold days and lepidopteran flight duration (Figure I). This pattern holds even after accounting for typical measures of accumulated heat. How general is this result across different regions of the planet and organisms, and how could this inform forecasting under environmental change?

Design and implementation

An Integrated Framework can move beyond independently leveraging Big Data and Experimental Frameworks. In an Integrated Framework, scaling across temperate regions of the Northern Hemisphere and taxa involves utilizing big data resources, including high-density professional monitoring and community science data, to examine co-located plant and insect phenological responses to climate variation at finer grain. *In situ* warming experiments across multiple sites [64] with differing climate context are chosen based on logistics and knowledge of initial big data results (Figure I). Experiments, including replicated field exclosures that limit species interactions and greenhouses or growth chamber-based experiments that manipulate variation in daily conditions, along with overall warming, provide a means to perturb the system under controlled conditions.

Analysis/interpretation

An Integrated Framework that combines the strengths of Big Data and Experimental Frameworks allows for strong inference of phenology change that can generalize over a broad extent [65]. In particular, workflows that dynamically assimilate experimental and big data results can be directly fed into process models for phenology responses that take into account overall warming and climatic extremes and that can make short- and longer-term forecasts (Figure I). Further experiments can be used to help reduce uncertainty in areas where models are particularly poor at prediction.

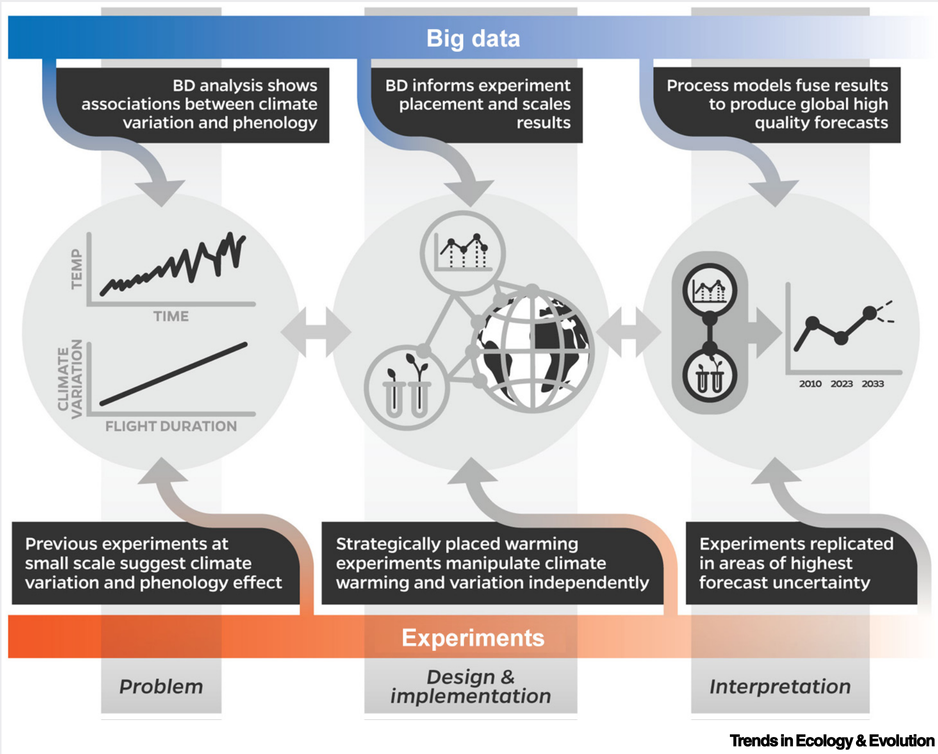


Figure I. A worked example to understand and predict changes in phenology using big data and experimental approaches that inform each stage of an Integrated Framework, from defining the problem (Problem) to study design (Design and Implementation), and interpretation of findings (Interpretation). This approach has feedbacks between interpretation and design that hasten new knowledge generation and applied outcomes.

for the placement of manipulations [29,30], improving the relevance of manipulations for predictions across scales and providing a better foundation for reducing spatial and temporal biases when combined with big data. For instance, to understand the response of plant communities to climate change, treatments in climate warming experiments should, but currently do not, match future climate predictions from big data [6]. Similarly, conclusions from the application of the Big Data Framework often include concerns that data are spatially biased and may not capture key environmental changes [31,32]. The Big Data Framework can then utilize experimental outcomes to better identify more targeted data streams, in terms of locations for data needs or data types relevant to causal processes (see worked examples in Boxes 2 and 3). Furthermore, the replication of experiments can be increased at low cost to an individual researcher by using big data from community science or Earth monitoring tools across larger areas [33]. Implementation of the Big Data Framework can also gain efficiency by narrowing the search for relevant model formulations that capture processes revealed through experimental manipulations. Finally, design integration and implementation can facilitate **data integration** from Big Data and Experimental Frameworks, resulting in **hybrid datasets** that can be analyzed to produce more reliable conclusions [1,17].

Analysis and modeling

Currently, the most extensive and successful efforts to leverage the strengths of these frameworks have focused on the analysis portion of the scientific process [17]. The development and

Box 3. Example of an Integrated Framework addressing insect conservation

Problem

Recently we have seen dramatic declines in the diversity and abundance of insects [66]; however, the scope, scale, and causes remain poorly understood. Big data approaches that feed vast opportunistic data streams into robust modeling frameworks could inform understanding the changes in occurrence for some clades and regions, such as North American bees [67]. However, these big data approaches are typically sparser in the past, limiting trends to coarse spatial scales. In addition, the stressors causing insect declines, which can be identified by experiments [11], are likely operating at finer scales.

Design and implementation

An Integrated Framework can be used to understand the magnitude and trends of potential stressors across clades and regions, while localizing stressor mechanisms. For example, Soroye *et al.* [67] collated broad-scale opportunistic data on bumble bees across North America and Europe, finding declines were driven by temperature variation. Despite these associations, the causality of temperature changes driving trends remains unknown, and other stressors, for example, insecticides, invasive species, and habitat loss, may also be influential [9,66]. Experiments manipulating temperatures, where treatments reflect observed changes from Soroye *et al.* [67], could estimate relevant effect sizes of temperature on population limitations, either as simple treatments within areas of known changes, as coordinated experiments across regions, or as factorial experiments that contrast temperature to other hypothesized stressors [11]. Experiments can then help identify gaps in insects' responses to stressors, which can be filled by big data streams utilizing technological advancements (e.g., computer vision, radar, acoustics, molecular techniques) to improve spatial, temporal, and taxonomic coverage of trends [68].

Analysis/interpretation

Experimental results can be linked with big data models in two ways. First, experimental results can alter covariates or the model structure in analysis, by identifying hidden interactions or important stressors. Second, experimental effect sizes could be used as priors in models or through joint likelihoods, updated as more trend data becomes available. Such priors could be parameterized to further hone inferences by drawing strength across species that are closely related or share similar traits. A challenge will be relating fine-scale experimental results to coarse-grained big data. While methods are nascent, one approach is scale-invariant model parameterizations [69], such as point process model formulations that use integrated intensity functions to scale from smaller to larger grains [70]. The outcome of this Integrated Framework can improve knowledge of mechanisms and scales at which stressors operate and ultimately inform interventions that maximize conservation outcomes.

use of models to link data generated from different data sources and inferential frameworks is growing rapidly [18,34,35]. These model integration tools can address some shortcomings in both frameworks. For example, using statistical models that incorporate experimental and systematically sampled data arising from **probabilistic samples** with the unplanned data streams from **nonprobabilistic samples**, common in big data, can increase precision and reduce biases [17,18]. Model integration methods such as hierarchical models can also improve the relevance of both Experimental and Big Data Frameworks [17,36]. For instance, Talluto *et al.* [37] illustrated how experiments on plant population growth at a fine spatial scale can be used as priors in a hierarchical Bayesian model to capture coarse-scale information on fundamental niches to predict species distributions. Taken together, model integration and analysis are key components of a broader Integrated Framework.

Interpretation and scope of inference

Finally, the interpretation of studies and the knowledge gained will be improved through unification. Linking experiments with big data will help break down the uncertainty that currently constrain each framework in isolation (Box 1). Inferences from the Big Data Framework often result in correlative associations when causality is desired. Advances in **causal modeling** and **counterfactual modeling** with big data provide an exciting way forward [38,39]. Yet, without the benchmark of causal experimental outcomes to resolve conflicting conclusions, the scope of causality obtained from such models remains unclear. In addition, the reliability, context, and applicability of experiments to real-world problems can be improved via integration with the Big Data Framework.

Problems the Integrated Framework can solve

An Integrated Framework could be applied to a wide range of ecology and conservation problems. We provide two examples of how this framework can be operationalized (Boxes 2 and 3) for different applications and outcome values for example, forecasting and delimiting spatial scale at which stressors operate. In the following section, we highlight four problems where the application of an Integrated Framework could be particularly fruitful.

Forecasting future state changes

Many forecasts utilize space-for-time substitutions, but these rely on assumptions of stationarity and are often poor at extrapolation to future states [40]. Forecasting is a key area for the integration of Experimental and Big Data Frameworks, but critically this integration should be bi-directional such that big data-driven forecasting models predict outcomes from experimental systems that perturb a system beyond its current state. Feedbacks between experimental results and forecast projections can not only lower uncertainties in those forecasts but completely shift parameters of underlying process models by correctly informing their mechanistic basis. To properly understand and predict the system, it is important that big data model results feed into design (e.g., magnitude and placement of manipulations) and data collection (e.g., earth monitoring tools, sensors, community science) of experiments used for forecasting. Box 2 provides a worked example of phenology forecasting integrating big data and experimental designs.

Ecological novelty and no-analog futures

Many problems in ecology and conservation either emerge from, or will generate, ecological novelty. Novelty can arise through climate conditions or the generation of new ecosystems, habitats, or communities that are unlike the ecological or evolutionary reference conditions [41]. Novelty is a challenging issue to address, as there is often limited information on how species respond to new conditions. Faske *et al.* [42] used transplant experiments for an invasive forest pest, the spongy moth (*Lymantria dispar* L.), to interpret factors that may limit range expansion. Experiments like these can be used with model integration, which may provide more reliable predictions into novel conditions than models parameterized using only existing range data. Similarly, Alexander *et al.* [43] argue that experiments are critical for understanding no-analog climate futures that are predicted by investigations using the Big Data Framework [44].

Scaling

The problem of translating patterns and processes across scales remains a fundamental challenge for ecology and conservation. An Integrated Framework provides a foundation for design, data, and model integration, which can reduce bias and lead to greater precision in conclusions. Change-of-support procedures that combine data with mismatched spatiotemporal extents are a key area in statistics that aims to address this issue [45]. For example, hierarchical and point-process models can be formulated in some situations to combine different sources of information while properly accounting for uncertainty [19]. These models have typically been applied in the context of linking different types of observational data (e.g., unstructured and structured observational data) measured at different scales [18]; however, there is potential for linking experiments and big data in this way to generate more reliable predictions across scales. Box 3 provides a worked example of how experiments and Big Data can inform scale of processes for insect declines.

Conservation action

Many management-based conservation actions involve perturbations, such as altering disturbance regimes, restoring habitats, or managing hydrology. Experiments can inform the reliability of perturbations, but by linking with big data, we can speed the implementation, expand the scope of investigations, and generalize understanding [46]. Furthermore, the Big Data

Framework can hone designs and implementation to refine perturbation insights, thereby leading to more effective interventions [46]. Box 3 provides an example of how an Integrated Framework can inform the location and types of interventions needed to mitigate the decline of insects.

Realizing an Integrated Framework

While we have shown nascent examples of integration at different phases of the scientific process, realizing the full potential of an Integrated Framework requires at least four fundamental changes. First, scientists and practitioners will need better transparency and objectivity in interpreting the scope and potential of different frameworks. We believe that some of the most practical approaches to breaking barriers and integrating Big Data and Experimental Frameworks are to focus on commonalities in vision and the need for clear, tangible benefits for ecology and conservation. Second, identifying opportunities to integrate these frameworks will benefit from training and better access to resources. We recommend integrating ecological problem solving, team teaching by big data and experimental scientists, and exposure to Big Data and Experimental Frameworks throughout the undergraduate and graduate curricula of ecology, conservation, and natural resource programs. Third, investment from both scientists and stakeholders is required. More specifically, research positions and funding models will need to prioritize and facilitate networks of researchers working on common goals over individuals and small teams [47]. Fourth, to remove inefficiencies in the flow of information (e.g., publications, data repositories) to and from Big Data and Experimental Frameworks, information transfer needs to be hardened into systems designs where experimentalists and modelers get relevant results but still quickly cross-communicate in an inclusive and readily accessible way.

Concluding remarks

The advent of the Big Data Framework has expanded the scope and scale of ecological problems that are solvable. Yet an Integrated Framework offers substantial opportunities for leveraging the Big Data and Experimental Frameworks to gain rapid and reliable answers to the conservation crisis. Although integration may not be needed for all ecological and conservation problems, we expect that many issues will benefit from using an Integrated Framework. There are several key questions that need answers regarding the implementation of an Integrated Framework, including when, where, and how to best apply these principles (see Outstanding questions). Yet, given the tremendous success of recent developments in model-based integration, we expect that integration throughout the scientific process will be an equally worthwhile pursuit.

Outstanding questions

Will the integration of experiments with big data throughout the scientific process lead to more reliable and timely conservation evidence and actions?

What is the future and importance of experiments, big data, and their integration, for interpreting causality?

Under what conditions does an Integrated Framework lead to differing conclusions, and how do scientists reconcile such conflict?

What types of technical (software, cyberinfrastructure) and physical (experimental sites and networks) infrastructure are needed to hasten knowledge integration?

What are the 'best practices' for implementing and evaluating each step of the scientific process when implementing an Integrated Framework?

Acknowledgments

Funding from the National Science Foundation (ISE-1952393) helped facilitate this project.

Declaration of interests

No interests are declared.

References

1. Niu, S.L. *et al.* (2020) Integrative ecology in the era of big data-From observation to prediction. *Sci. China Earth Sci.* 63, 1429–1442
2. Farley, S.S. *et al.* (2018) Situating ecology as a big-data science: current advances, challenges, and solutions. *Bioscience* 68, 563–576
3. Boyd, J.R. *et al.* (2023) We need to talk about nonprobability samples. *Trends Ecol. Evol.* 38, 513–531
4. Worthington, T.A. *et al.* (2020) Harnessing big data to support the conservation and rehabilitation of mangrove forests globally. *One Earth* 2, 429–443
5. Betts, M.G. *et al.* (2021) When are hypotheses useful in ecology and evolution? *Ecol. Evol.* 11, 5762–5776
6. Korell, L. *et al.* (2020) We need more realistic climate change experiments for understanding ecosystems of the future. *Glob. Chang. Biol.* 26, 325–327
7. Ockendon, N. *et al.* (2021) Effectively integrating experiments into conservation practice. *Ecol. Solut. Evid.* 2, e12069
8. Wiersma, Y.F. (2022) *Experimental Landscape Ecology*, Springer International Publishing
9. Paine, R.T. (1966) Food web complexity and species diversity. *Am. Nat.* 100, 65–75
10. Jochum, M. *et al.* (2020) The results of biodiversity-ecosystem functioning experiments are realistic. *Nat. Ecol. Evol.* 4, 1485–1494
11. Boyes, D.H. *et al.* (2021) Is light pollution driving moth population declines? A review of causal mechanisms across the life cycle. *Insect Conserv. Diver.* 14, 167–187
12. Henry, H.A.L. *et al.* (2018) Increased soil frost versus summer drought as drivers of plant biomass responses to reduced precipitation: results from a globally coordinated field experiment. *Ecosystems* 21, 1432–1444

13. Weisser, W. *et al.* (2023) Experiments are needed to quantify the main causes of insect decline. *Biol. Lett.* 19, 20220500
14. Currie, D.J. (2019) Where Newton might have taken ecology. *Glob. Ecol. Biogeogr.* 28, 18–27
15. Nathan, R. *et al.* (2022) Big-data approaches lead to an increased understanding of the ecology of animal movement. *Science* 375, eabg1780
16. Tuia, D. *et al.* (2022) Perspectives in machine learning for wildlife conservation. *Nat. Commun.* 13, 792
17. Benedetti-Cecchi, L. *et al.* (2018) Hybrid datasets: integrating observations with experiments in the era of macroecology and big data. *Ecology* 99, 2654–2666
18. Fletcher, R.J. *et al.* (2019) A practical guide for combining data to model species distributions. *Ecology* 100, e02710
19. Isaac, N.J.B. *et al.* (2020) Data integration for large-scale models of species distributions. *Trends Ecol. Evol.* 35, 56–67
20. Jaureguiberry, P. *et al.* (2022) The direct drivers of recent global anthropogenic biodiversity loss. *Sci. Adv.* 8, eabm9982
21. Thierry, M. *et al.* (2022) The presence of multiple parasitoids decreases host survival under warming, but parasitoid performance also decreases. *Proc. R. Soc. B* 289, 20220121
22. Hurlbert, S.H. (1984) Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.* 54, 187–211
23. Durden, J.M. *et al.* (2017) Integrating “big data” into aquatic ecology: challenges and opportunities. *Limnol. Oceanogr. Bull.* 26, 101–108
24. Prosser, J.I. (2020) Putting science back into microbial ecology: a question of approach. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 375, 20190240
25. Peters, D.P.C. *et al.* (2018) An integrated view of complex landscapes: a big data-model integration approach to transdisciplinary science. *Bioscience* 68, 653–669
26. Preisser, E.L. *et al.* (2005) Scared to death? The effects of intimidation and consumption in predator–prey interactions. *Ecology* 86, 501–509
27. Creel, S. *et al.* (2013) Underestimating the frequency, strength and cost of antipredator responses with data from GPS collars: an example with wolves and elk. *Ecol. Evol.* 3, 5189–5200
28. Suraci, J.P. *et al.* (2019) Fear of humans as apex predators has landscape-scale impacts from mountain lions to mice. *Ecol. Lett.* 22, 1578–1586
29. Gilby, B. *et al.* (2018) Spatial restoration ecology: placing restoration in a landscape context. *Bioscience* 68, 1007–1019
30. Rosenman, E.T.R. and Owen, A.B. (2021) Designing experiments informed by observational studies. *J. Causal Infer.* 9, 147–171
31. Engemann, K. *et al.* (2015) Limited sampling hampers "big data" estimation of species richness in a tropical biodiversity hotspot. *Ecol. Evol.* 5, 807–820
32. Hughes, A.C. *et al.* (2021) Sampling biases shape our view of the natural world. *Ecography* 44, 1259–1269
33. Fisher, A.G. *et al.* (2021) Remote sensing of trophic cascades: multi-temporal landsat imagery reveals vegetation change driven by the removal of an apex predator. *Landsc. Ecol.* 36, 1341–1358
34. Isaac, N.J.B. *et al.* (2014) Statistics for citizen science: extracting signals of change from noisy ecological data. *Methods Ecol. Evol.* 5, 1052–1060
35. Zipkin, E.F. and Saunders, S.P. (2018) Synthesizing multiple data types for biological conservation using integrated population models. *Biol. Conserv.* 217, 240–250
36. Zhou, X.K. *et al.* (2021) Deep correlation mining based on hierarchical hybrid networks for heterogeneous big data recommendations. *IEEE Trans. Comput. Soc. Syst.* 8, 171–178
37. Talluto, M.V. *et al.* (2016) Cross-scale integration of knowledge for predicting species ranges: a metamodelling framework. *Glob. Ecol. Biogeogr.* 25, 238–249
38. Raita, Y. *et al.* (2021) Big data, data science, and causal inference: a primer for clinicians. *Front. Med.* 8, 678047
39. Laubach, Z.M. *et al.* (2021) A biologist's guide to model selection and causal inference. *Proc. R. Soc. B* 288, 20202815
40. Damgaard, C. (2019) A critique of the space-for-time substitution practice in community ecology. *Trends Ecol. Evol.* 34, 416–421
41. Heger, T. *et al.* (2019) Towards an integrative, eco-evolutionary understanding of ecological novelty: studying and communicating interlinked effects of global change. *Bioscience* 69, 888–899
42. Faske, T.M. *et al.* (2019) Can gypsy moth stand the heat? A reciprocal transplant experiment with an invasive forest pest across its southern range margin. *Biol. Invasions* 21, 1365–1378
43. Alexander, J.M. *et al.* (2016) When climate reshuffles competitors: a call for experimental macroecology. *Trends Ecol. Evol.* 31, 831–841
44. Fitzpatrick, M.C. and Dunn, R.R. (2019) Contemporary climatic analogs for 540 North American urban areas in the late 21st century. *Nat. Commun.* 10, 614
45. Zipkin, E.F. *et al.* (2021) Addressing data integration challenges to link ecological processes across scales. *Front. Ecol. Environ.* 19, 30–37
46. Shiklomanov, A.N. *et al.* (2019) Enhancing global change experiments through integration of remote-sensing techniques. *Front. Ecol. Environ.* 17, 215–224
47. Orlikowska, E.H. *et al.* (2016) Gaps in ecological research on the world's largest internationally coordinated network of protected areas: a review of Natura 2000. *Biol. Conserv.* 200, 216–227
48. Kitchin, R. (2014) Big Data, new epistemologies and paradigm shifts. *Big Data Soc.* 1, 2053951714528481
49. Wagner, D.L. *et al.* (2021) A window to the world of global insect declines: moth biodiversity trends are complex and heterogeneous. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2002549117
50. Rosenberg, K.V. *et al.* (2019) Decline of the North American avifauna. *Science* 366, 120–124
51. Brodie, J.F. *et al.* (2021) The decline of mammal functional and evolutionary diversity worldwide. *Proc. Natl. Acad. Sci. U. S. A.* 118, e1921849118
52. Metcalf, A.L. *et al.* (2019) Microtargeting for conservation. *Conserv. Biol.* 33, 1141–1150
53. Giuliani, G. *et al.* (2020) Knowledge generation using satellite earth observations to support sustainable development goals (SDG): a use case on land degradation. *Int. J. Appl. Earth Obs. Geoinf.* 88, 102068
54. Xia, J.Y. *et al.* (2020) Research challenges and opportunities for using big data in global change biology. *Glob. Chang. Biol.* 26, 6040–6061
55. Serra-Diaz, J.M. *et al.* (2018) Big data of tree species distributions: how big and how good? *For. Ecosyst.* 4, 30
56. Dutta, R. *et al.* (2016) Big data integration shows Australian bush-fire frequency is increasing significantly. *R. Soc. Open Sci.* 3, 150241
57. Andersen, A.N. (2021) Faunal responses to fire in Australian tropical savannas: insights from field experiments and their lessons for conservation management. *Divers. Distrib.* 27, 828–843
58. Loft, L. *et al.* (2019) Effectiveness and equity of payments for ecosystem services: real-effort experiments with Vietnamese land users. *Land Use Policy* 86, 218–228
59. Vredenburg, V.T. (2004) Reversing introduced species effects: experimental removal of introduced fish leads to rapid recovery of a declining frog. *Proc. Natl. Acad. Sci. U. S. A.* 101, 7646–7650
60. McCleery, R.A. *et al.* (2015) Marsh rabbit mortalities tie pythons to the precipitous decline of mammals in the Everglades. *Proc. R. Soc. B* 282, 20150120
61. Franklin, J. *et al.* (2017) Big data for forecasting the impacts of global change on plant communities. *Glob. Ecol. Biogeogr.* 26, 6–17
62. Morisette, J.T. *et al.* (2009) Tracking the rhythm of the seasons in the face of global change: phenological research in the 21st century. *Front. Ecol. Environ.* 7, 253–260
63. Liu, L.L. and Zhang, X.Y. (2020) Effects of temperature variability and extremes on spring phenology across the contiguous United States from 1982 to 2016. *Sci. Rep.* 10, 17952
64. Collins, C.G. *et al.* (2021) Experimental warming differentially affects vegetative and reproductive phenology of tundra plants. *Nat. Commun.* 12, 3442
65. Hanninen, H. *et al.* (2019) Experiments are necessary in process-based tree phenology modelling. *Trends Plant Sci.* 24, 199–209
66. Wagner, D.L. (2020) Insect declines in the Anthropocene. *Annu. Rev. Entomol.* 65, 457–480

67. Soroye, P. *et al.* (2020) Climate change contributes to widespread declines among bumble bees across continents. *Science* 367, 685–688
68. van Klink, R. *et al.* (2022) Emerging technologies revolutionise insect ecology and monitoring. *Trends Ecol. Evol.* 37, 872–885
69. Dorazio, R.M. (2014) Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. *Glob. Ecol. Biogeogr.* 23, 1472–1484
70. Fletcher Jr., R.J. *et al.* (2023) Addressing the problem of scale that emerges with habitat fragmentation. *Glob. Ecol. Biogeogr.* 32, 828–841