MEETING REVIEWS

Minimizing Data Waste: Conservation in the Big Data Era

Allison D. Binley[1] , Brandon P. M. Edwards[1] , Gabriel Dansereau[2] , Elly C. Knight[3] , and Iman Momeni-Dehaghi[1]

[1]Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario K1S 5B6 Canada

[2]Department of Biology, University of Montreal, 2900 Edouard Montpetit Blvd, Montreal, Quebec H3T 1J4 Canada

[3]Alberta Biodiversity Monitoring Institute, University of Alberta, 116 St & 85 Ave, Edmonton, Alberta T6G 2R3 Canada

Session motivation and objectives

Advances in computing, statistics, and technology over the past few decades has resulted in the accumulation of massive amounts of biodiversity data, as well as novel methods for using and integrating them (Miller et al. 2019). Data that have been collected for decades or even centuries can now be analyzed and applied in brand new ways. These data include alternative sources of information such as citizen science (also called community science) programs (Butcher et al. 1990, Sullivan et al. 2014, Hudson et al. 2017). In the midst of a global biodiversity crisis, these databases may hold the key to detecting important obstacles and threats to conservation, as well as determining the best interventions before it is too late. The ecological processes in question frequently occur across broad spatial and temporal scales; migratory species' ranges can span thousands of kilometers, and the impacts of threats such as climate change cannot necessarily be assessed on a local scale. Conventional methods of biodiversity monitoring may therefore be inadequate in the face of broad scale change. The individual researcher can generally only collect data on a localized scale, yet with the proper data practices, this information can contribute to a better understanding of the bigger picture (Sutherland et al. 2009).

Data collected through biodiversity monitoring are a crucial component of solving any conservation problem. Gaining information about the state of an ecological system can inform better decisions to be made surrounding protections of that system (Bennett et al. 2018). However, with biodiversity rapidly declining, monitoring can delay much-needed action and take up valuable and limited resources, and therefore may not always be the best option (Bennett et al. 2018, Buxton et al. 2020). Additionally, in many cases the data needed to answer pressing conservation questions already exist, and need only to be made more accessible or simply used. The increasing availability of open data through large data repositories and alternative sources of data, such as citizen science programs, means that researchers should not have to compromise between fast action and informed action. Statistical methods of data integration can allow researchers to fill perceived spatial and temporal knowledge gaps using multiple existing datasets (Miller et al. 2019, Zipkin et al. 2019, Isaac et al. 2020), and analytical tools are available that can help quantify whether more data are needed (Canessa et al. 2015, Bennett et al. 2018).

In the symposium entitled "Minimizing Data Waste: Conservation in the Big Data Era," we explored how using open and available data not only makes the best and most efficient use of limited resources, but can lead to better conservation outcomes. We investigated how data integration helps improve our understanding of species trends and distributions, better evaluate ecological systems, and redistribute limited resources from monitoring to action. We emphasized why these advances in open data and data integration are critical not only for minimizing data waste, but also conducting better conservation research and directly improving conservation management decisions.

Following are summaries of each of the four speakers' presentations. The first presentation (Knight) provided a definition of data integration, and showed several examples of how integrating multiple datasets led to improve knowledge of common nighthawks (*Chordeiles minor*) across their full annual cycle. The second presentation (Dansereau) demonstrated how using large open datasets helped to reveal key insights into the distribution of ecological uniqueness, and identify potential conservation targets over broad spatial scales. The third presentation (Momeni-Dehaghi) showed how available citizen science data for Monarch butterflies (*Danaus plexippus*) was used to map premigration distributions, with implications toward their mortality rates. The final presentation (Binley) sought to compare the outcomes of using freely available citizen science data versus paid professional surveys to prioritize conservation action, both in terms of costs and biodiversity. The final section of this report synthesizes these presentations, addresses some topics that were covered in the Q&A roundtable sessions, and provides an outlook on the future of conservation in the big data era. Key points from this symposium are summarized in Fig. 1.

Presentation summaries

Elly Knight. Full annual cycle conservation of a declining, nocturnal bird: A data integration journey

This presentation demonstrated how data integration of existing monitoring datasets can improve our understanding of ecology and conservation needs. Data integration was defined broadly as the statistical combination of datasets collected under varying protocols (Miller et al. 2019, Isaac et al. 2020). Integration can be particularly useful for answering questions across large temporal and/or spatial extents and for poorly understood species because it can pool information from multiple sparse datasets. Many data integration approaches are centered around accounting for differences in probability of detection and its subcomponents, availability (the probability an individual provides a cue for detection if it is
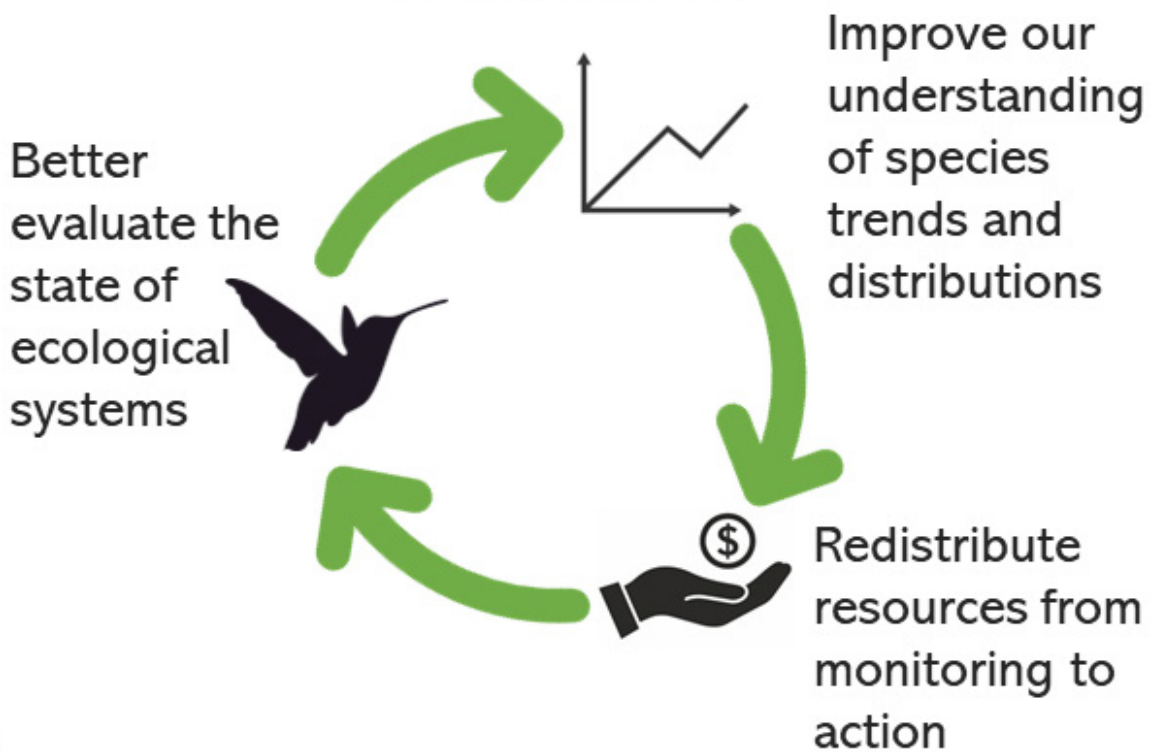
Fig. 1. Infographic summarizing the key points from the symposium "Minimizing Data Waste: Conservation in the Big Data Era". By making use of the most amount of open and available data as possible, researchers, government agencies, and conservation practitioners can better understand the trends and distributions of species of interest, which can lead us to better understand and evaluate the state of the current ecological system. By having better knowledge on the system, conservation managers can make better use of limited budgets by spending more of the budget on conservation action, rather than just monitoring alone.

present; sensu Marsh and Sinclair 1989) and perceptibility (the probability that cue is detected if the cue is provided) by using submodel(s) to estimate the probability of detection (e.g. MacKenzie et al. 2002, Sólymos et al. 2013, Miller et al. 2019).

Knight presented several data integration examples for the common nighthawk, which is a declining, poorly understood, nocturnal, long-distance migratory bird whose distribution spans most of the western hemisphere (Brigham et al. 2020). First, by combining general citizen science datasets with more targeted datasets for understanding population trends and habitat relationships (Knight et al. 2021*b*). Next, by integrating various human observer datasets to calculate the probability of common nighthawk detection under various conditions (Sólymos et al. 2013, Edwards et al. 2022). Finally, by using those probabilities to integrate multiple monitoring datasets and evaluate various full annual cycle hypotheses for causes of common nighthawk population declines. The full annual cycle analysis was informed by satellite tracking data (Knight et al. 2021*a*).

Data integration provided multiple benefits along the journey to understand common nighthawk population declines. First, integrating general monitoring datasets with targeted nocturnal monitoring increased the probability of detecting a 30% population decline from 38% to 69% (Knight et al. 2021*b*). Integrating the two monitoring programs also improved the predictive performance of species distribution modeling. Data integration for calculating probability of detection improved the temporal, spatial, and data type coverage of statistical offsets. Finally, data integration for understanding full annual cycle causes of population declines resulted in more significant population trends detected and stronger results more aligned with common nighthawk ecology. Together, this data integration journey emphasizes the value of data integration for overcoming statistical hurdles, particularly for conservation research of data sparse species.

In addition to large citizen science datasets used in these analyses like eBird and the North American Breeding Bird Survey, the analyses also used the Boreal Avian Modelling project database (Barker et al. 2015), the Canadian Nightjar Survey dataset (Knight et al. 2019), the Nightjar Survey Network dataset (Centre for Conservation Biology 2022), and the Common Nighthawk Migratory Connectivity Project (Knight et al. 2021*a*). Those datasets were accessed via several data portals which also provide open access to many other citizen science and public datasets (Table 1). The data integration approaches used were based on the "QPAD" framework (Sólymos et al. 2013), which is a flexible framework for data integration that allows the user to estimate availability and perceptibility and use those estimates as statistical offsets in a wide range of modeling approaches. Models and estimates for over 100 boreal landbird species are available in the QPAD R package.

Gabriel Dansereau. Extending ecological uniqueness indicators to broader spatial extents using open data

Open data hold an underexploited potential to improve ecological indicators used in conservation research. For instance, indicators such as ecological uniqueness, which can help identify areas with exceptional species composition and potential conservation targets, are often limited to local or regional scales because of sampling limitations for single studies. In contrast, this presentation showed how open data allows measuring uniqueness over broad spatial extents and new data types while improving knowledge about the indicator itself.

Based on Dansereau et al. (2022), this presentation showed that uniqueness could be predicted over broad spatial extents, revealing additional sources of variation in space. They used species distribution modeling to predict community composition across North America based on eBird citizen science data

Table 1. Some data acquisition tools that were used in the presenters' analyses.

Tool	Description	Link or Reference
WildTrax	Online portal and R package for managing, storing, sharing, processing, and discovering biological and environmental sensor data, including data from autonomous recording units and wildlife cameras.	https://www.wildtrax.ca
NatureCounts	Collect, archive, interpret, and access wildlife data to advance the understanding of bird populations across the Western Hemisphere	https://www.naturecounts.ca
Movebank	A free, online database of animal tracking data that helps animal tracking researchers to manage, share, protect, analyze, and archive their data.	Kays et al. (2022)
QPAD	An R package to access statistical offsets to account for detectability for over 100 species of breeding boreal birds.	Sólymos et al. (2013)
bbsBayes	An R package to allow users to download, stratify, analyze, and interpret data from the North American Breeding Bird Survey	Edwards and Smith (2021)
auk	An R package to allow users to access millions of bird observations from the citizen science platform eBird	Strimas-Mackey et al. (2018)
Ebirdst	An R package to allow users to explore status and trend estimates of hundreds of species of birds	Strimas-Mackey et al. (2021)

Note: Owners and managers of large biological datasets can facilitate use of the dataset by creating and making available free and easy-to-use software packages and database interfaces such as the ones listed here.

(Sullivan et al. 2014). Then, they measured ecological uniqueness using the local contributions to beta diversity (Legendre and De Cáceres 2013) and explored their spatial variation across different regions and scales. Uniqueness and its relationship to species richness changed according to the area under study and was affected by regional factors such as extent size, richness profile, and proportion of rare species. As a result, sites identified as unique may vary according to regional characteristics, which should be considered when this indicator is used for conservation recommendations.

In addition, this presentation showed how open data could help extend uniqueness assessments to species interactions. They combined community predictions with an open interaction metaweb (Strydom et al. 2022) to produce localized predictions of ecological networks, then measured uniqueness separately based on interaction and community composition (Poisot et al. 2017). Interaction uniqueness showed a different spatial distribution from community uniqueness over whole regions, highlighting that sites and areas may be unique in one community aspect and not the other (e.g. unique communities without unique interaction networks). Therefore, considering different community components through open data sources can reveal additional important conservation targets, especially over broad spatial scales.

Iman Momeni-Dehaghi. Mapping the premigration distribution of eastern Monarch butterflies using community science data

To conserve migratory species effectively, we need to know their distribution at different stages of their life cycle. For the Monarch butterfly (*Danaus plexippus*), although it is well-known that they

overwinter in oyamel fir (*Abies religiosa*) forests in the mountains west of Mexico City, the natal origins of these overwintering Monarchs are the only information we have about their premigration distribution (e.g. Flockhart et al. 2017). However, the premigration distribution and the natal origins of overwintering Monarchs (i.e. postmigration data) can be considered equivalent only if we assume that Monarchs originated from different regions have a similar mortality rate during their migration (Momeni-Dehaghi et al. 2021).

In his talk, Momeni-Dehaghi demonstrated how the premigration map they developed using community science data can contribute to the conservation of Monarch butterflies (Momeni-Dehaghi et al. 2021). To estimate Monarchs' premigration distribution, the authors used data reported by citizen scientists in the Journey North program (http://www.journeynorth.org/) before Monarchs start their fall migration (i.e. before migration mortality), controlling for sampling bias. Momeni-Dehaghi compared the resulting distribution to distributions estimated using postmigration data (i.e. isotopic-based natal origin assignments) to determine whether migration mortality varies between butterflies originating from different regions. This comparison suggests that Monarchs starting their migration from North-central breeding region have a higher mortality rate than other regions. In contrast, those which originated from Northwest and Southeast breeding regions have a lower mortality rate relative to other Monarchs. Their premigration distribution map will be useful in future studies estimating the rates, distribution, and causes of mortality in migrating Monarchs. These results have clear implications for when and where conservation action should be prioritized for protecting these species. Given the broad, spatially complex ranges of migratory species such as the Monarch butterfly, large open datasets such as those collected through community science programs are essential to understanding the threats that face them throughout their full annual cycle.

Allison Binley. Redistributing conservation resources from monitoring to action using citizen science

An additional benefit of using freely available open data is that it serves as an effective means to collect vast quantities of data in a cost-efficient manner. This may prove invaluable to conservation efforts, as using crowd-sourced data to inform decisions can allow managers to redirect limited funds towards action rather than monitoring. This presentation demonstrated the quantitative benefit of using citizen science data for setting conservation priorities in an applied conservation setting.

Using data from the BirdReturns conservation program implemented by The Nature Conservancy in central California, rice farms were prioritized for conservation action based on the modeled probability of detection of seven shorebird species using two datasets: eBird citizen science data, and monitoring data collected through professional surveys conducted by The Nature Conservancy (Robinson et al. 2020). The value of these prioritizations was assessed using an integrated dataset that combined both the community and professionally collected data as a benchmark. Prioritizations conducted using the professional monitoring data were subject to a monitoring penalty, where the cost of monitoring was deducted from the overall budget, leaving less remaining to pay for conservation action. Prioritizations were then run across a range of different available budgets. The authors predicted that decisions based solely on eBird data would be preferable at lower budgets given that more money can be spent on action rather than monitoring, but that the more targeted professional monitoring may provide better value at larger budgets.

Contrary to those predictions, prioritizing detections across all seven species based on the model using eBird data resulted in the greatest overall value across all budgets. The difference was greatest at lower budgets, but prioritizations based on eBird data consistently performed better until the budget was large enough so that all properties could be enrolled in the program. Furthermore, prioritizations based on citizen science data performed comparably to those based on the integrated model (i.e. the best available information). Even when the monitoring penalty was removed from the professional monitoring prioritization, allowing for a more direct comparison of information content, the eBird prioritization performed comparably or better than professional monitoring. This demonstrates that, in this case study, eBird citizen science data matched or surpassed the capacity of professional monitoring data to inform conservation decisions.

This presentation quantified the trade-offs between monitoring and action, to better illustrate to conservation managers the potential risks associated with unnecessary data collection. In this case study, there was no benefit to spending money on professional monitoring at any budget. Using openly and freely available data, resources can be redistributed towards actions that will directly benefit biodiversity, ultimately resulting in better overall outcomes.

Synthesis

This symposium presented a small subset of advances in ecology and conservation through the use of Big Data. While collecting data will always remain important in these fields, many pressing questions can be more immediately and satisfactorily answered using data already available. The selected speakers and talks presented here demonstrated how innovative and integrative approaches using Big Data are a necessary next step in the evolution of the field of conservation, not only in making the most efficient use of resources, but also for better understanding the ecological systems we are trying to protect. Each presentation explicitly showed how the knowledge obtained from Big Data integration represents an improvement over using locally sourced data alone. The first three talks provided an overview on how open data integration can enhance our understanding of several biodiversity metrics that are critical to making conservation decisions, and the last quantified the direct benefit of doing so, both in terms of dollars and biodiversity.

During the roundtable discussions, one topic brought up was the well-known issue of taxonomic biases in the collection of monitoring data (Mair and Ruete 2016, Troudet et al. 2017, Binley et al. 2023). Indeed, this was evident in the selection of speakers for this symposium. Although all speakers presented different applications and analyses of big data in ecology, three of the four speakers focused on birds and analyses using large bird-focused datasets. Whereas Knight, Dansereau, and Binley had access to millions of bird observations from large citizen science platforms (i.e. eBird, BBS, etc.), Momeni-Dehaghi had access to only one program with much less available data. We agree that this limited availability of open and available data for other under-sampled taxa does pose a large issue in tackling broad-scale biodiversity issues across a range of taxa. For professionally collected data, we encourage researchers to continue making their data as open and available as possible. For example, programs such as entoGEM (Grames et al. 2022) provide a great example of a collection of insect data, and the website WildTrax (https://wildtrax.ca) provides access to data collected by an array of environmental sensors such as autonomous recording units and camera traps, as well as professional point counts. We also note the rise in availability of citizen science data through programs such as eButterfly (Prudic et al. 2017), as well as the use of citizen science data in studying jellyfish blooms (Marambio et al. 2021), fisheries management (Bellquist et al. 2022), amphibian

conservation (Lee et al. 2021), mosquito surveillance (Sousa et al. 2022), and for threatened species monitoring (Soroye et al. 2022). We encourage the continual development of these citizen science programs that target undersampled taxa in creative and engaging ways.

Another topic of discussion surrounded the availability of datasets and use of tools to access these large datasets. The speakers discussed some of the tools they used to access the data for their analyses (Table 1), but noted that there are likely several other tools beyond the scope of what was highlighted in these presentations. Given the increasing need to house large datasets in databases, the development of computational tools to access these data has become paramount, and we encourage owners and managers of large biological databases to continue developing open easy-to-use tools to allow for further access to these large databases (Fortunato and Galassi 2021). Additionally, we encourage the movement to continue to make data findable, accessible, interoperable, and reusability (i.e. FAIR data; Wilkinson et al. 2016) to further facilitate data integration in ecology and evolution (O'Dea et al. 2021).

Finally, a major topic during the roundtable discussion surrounded that of the use and reliability of citizen science data. One common criticism of citizen science data is the taxonomic and spatial biases, as well as potential for lack of (or differences in) structure in data collection. Binley noted, however, that though these are fair criticisms for citizen science data, these issues also commonly arise in professionally collected data as well. Binley also noted that many large-scale citizen science platforms such as eBird have thought extensively about controlling for these biases (Johnston et al. 2021), and so researchers that are hesitant about using alternative data sources should take some comfort in knowing that many of the expected biases have been accounted for (Ellwood et al. 2017).

Conservation implications

The objective of this symposium was to demonstrate examples of large, open biodiversity databases that already exist, as well as methods for working with and integrating these datasets to make the best use of available information. It is well known that we are in a global biodiversity crisis, and so we must act fast to understand the current states of species' ecologies, including population trends, distributions, excess causes of mortalities, etc., and use this knowledge to implement conservation action to species who need it the most. This is particularly important for modern conservation research, which continues to face limited conservation budgets yet allocate over half of the budget to just monitoring (Buxton et al. 2020). By making use of data already available, conservation managers and practitioners can look to implement conservation actions sooner rather than later, to avoid the situation of species being "monitored to death" (Lindenmayer et al. 2013).

The concepts explored during this session are of general interest and applicability to anyone in the field of ecology or conservation, particularly those conducting work on broad spatial and temporal scales. Given the vast investment of time and resources that has already gone into collecting these data, and the time-sensitive nature of conservation research, it is vital that researchers not only familiarize themselves with what data is already available, but also the methods required to make use of them. In a world where the volume of biodiversity data continues to rise, yet biodiversity itself continues to decline, the ability to make use of open and available datasets to inform future conservation decisions in a timely manner will play a critical role in preventing further extinctions.

Author contributions

A. Binley and B. Edwards contributed equally to this manuscript.

Literature Cited

Barker, N. K. S., P. C. Fontaine, S. G. Cumming, D. Stralberg, A. Westwood, E. M. Bayne, P. Sólymos, F. K. A. Schmiegelow, S. J. Song, and D. J. Rugg. 2015. Ecological monitoring through harmonizing existing data: lessons from the boreal avian modelling project: data management of the BAM project. Wildlife Society Bulletin 39:480–487.

Bellquist, L., W. J. Harford, F. Hurd, A. Jackson, J. D. Prince, J. Freiwald, A. Neumann, J. Likins, and J. R. Wilson. 2022. Use of management strategy evaluation to understand the value of citizen science in managing an iconic California recreational fishery. Estuarine, Coastal and Shelf Science 278:108112.

Bennett, J. R., S. L. Maxwell, A. E. Martin, I. Chadès, L. Fahrig, and B. Gilbert. 2018. When to monitor and when to act: value of information theory for multiple management units and limited budgets. Journal of Applied Ecology 55:2102–2113.

Binley, A. D., J. Vincent, T. Rytwinski, C. Proctor, E. Urness, S. Davis, P. Soroye, and J. R. Bennett. 2023. Patterns of community science data use in peer-reviewed research on biodiversity. Biological Conservation 280:1–10. https://doi.org/10.1016/j.biocon.2023.109985

Brigham, R. M., J. Ng, R. G. Poulin, and S. D. Grindal. 2020. Common nighthawk (*Chordeiles minor*). *In* S. M. Billerman, B. K. Keeney, P. G. Rodewald, and T. S. Schulenberg (Eds). Birds of the world. Cornell Lab of Ornithology, Ithaca, New York, USA.

Butcher, G. S., M. R. Fuller, L. S. McAllister, and P. H. Geissler. 1990. An evaluation of the Christmas bird count for monitoring population trends of selected species. Wildlife Society Bulletin (1973–2006) 18:129–134.

Buxton, R. T., S. Avery-Gomm, H.-Y. Lin, P. A. Smith, S. J. Cooke, and J. R. Bennett. 2020. Half of resources in threatened species conservation plans are allocated to research and monitoring. Nature Communications 11:4668.

Canessa, S., G. Guillera-Arroita, J. J. Lahoz-Monfort, D. M. Southwell, D. P. Armstrong, I. Chadès, R. C. Lacy, and S. J. Converse. 2015. When do we need more data? A primer on calculating the value of information for applied ecologists. Methods in Ecology and Evolution 6:1219–1228.

Centre for Conservation Biology. 2022. Nightjar survey network protocol. http://www.nightjars.org/

Dansereau, G., P. Legendre, and T. Poisot. 2022. Evaluating ecological uniqueness over broad spatial extents using species distribution modelling. Oikos 2022:1–13.

Edwards, B. P. M., and A. C. Smith. 2021. bbsBayes: an R package for hierarchical Bayesian analysis of north American breeding bird survey data. Journal of Open Research Software 9:17.

Edwards, B. P. M., et al. 2022. Point count offsets for estimating population sizes of North American Landbirds. Ibis. https://doi.org/10.1111/ibi.13169

Ellwood, E. R., T. M. Crimmins, and A. J. Miller-Rushing. 2017. Citizen science and conservation: recommendations for a rapidly moving field. Biological Conservation 208:1–4.

Flockhart, D. T. T., L. P. Brower, M. I. Ramirez, K. A. Hobson, L. I. Wassenaar, S. Altizer, and D. R. Norris. 2017. Regional climate on the breeding grounds predicts variation in the natal origin of monarch butterflies overwintering in Mexico over 38 years. Global Change Biology 23:2565–2576.

Fortunato, L., and M. Galassi. 2021. The case for free and open source software in research and scholarship. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 379. https://doi.org/10.1098/rsta.2020.0079

Grames, E. M., et al. 2022. A framework and case study to systematically identify long-term insect abundance and diversity datasets. Conservation Science and Practice 4:e12687.

Hudson, M.-A. R., C. M. Francis, K. J. Campbell, C. M. Downes, A. C. Smith, and K. L. Pardieck. 2017. The role of the north American breeding bird survey in conservation. The Condor 119:526–545.

Isaac, N. J. B., et al. 2020. Data integration for large-scale models of species distributions. Trends in Ecology & Evolution 35:56–67.

Johnston, A., W. M. Hochachka, M. E. Strimas-Mackey, V. Ruiz Gutierrez, O. J. Robinson, E. T. Miller, T. Auer, S. T. Kelling, and D. Fink. 2021. Analytical guidelines to increase the value of community science data: an example using eBird data to estimate species distributions. Diversity and Distributions 27:1265–1277.

Kays, R., et al. 2022. The Movebank system for studying global animal movement and demography. Methods in Ecology and Evolution 13:419–431.

Knight, E., et al. 2019. Canadian Nightjar Survey Protocol. http://wildresearch.ca/wp-content/uploads/2019/05/National-Nightjar-Survey-Protocol-WildResearch-2019.pdf

Knight, E. C., et al. 2021a. Comprehensive estimation of spatial and temporal migratory connectivity across the annual cycle to direct conservation efforts. Ecography 44:665–679.

Knight, E. C., A. C. Smith, R. M. Brigham, and E. M. Bayne. 2021b. Combination of targeted monitoring and breeding bird survey data improves population trend estimation and species distribution modeling for the common nighthawk. Ornithological Applications 123:duab005.

Lee, T. S., N. L. Kahal, H. L. Kinas, L. A. Randall, T. M. Baker, V. A. Carney, K. Kendell, K. Sanderson, and D. Duke. 2021. Advancing amphibian conservation through citizen science in urban municipalities. Diversity 13:211.

Legendre, P., and M. De Cáceres. 2013. Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. Ecology Letters 16:951–963.

Lindenmayer, D. B., M. P. Piggott, and B. A. Wintle. 2013. Counting the books while the library burns: why conservation monitoring programs need a plan for action. Frontiers in Ecology and the Environment 11:549–555.

MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. Andrew Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83:2248–2255.

Mair, L., and A. Ruete. 2016. Explaining spatial variation in the recording effort of citizen science data across multiple taxa. PLoS One 11:e0147796.

Marambio, M., et al. 2021. Unfolding jellyfish bloom dynamics along the Mediterranean basin by transnational citizen science initiatives. Diversity 13:274.

Marsh, H., and D. F. Sinclair. 1989. Correcting for visibility bias in strip transect aerial surveys of aquatic Fauna. The Journal of Wildlife Management 53:1017–1024.

Miller, D. A. W., K. Pacifici, J. S. Sanderlin, and B. J. Reich. 2019. The recent past and promising future for data integration methods to estimate species' distributions. Methods in Ecology and Evolution 10:22–37.

Momeni-Dehaghi, I., J. R. Bennett, G. W. Mitchell, T. Rytwinski, and L. Fahrig. 2021. Mapping the premigration distribution of eastern monarch butterflies using community science data. Ecology and Evolution 11:11275–11281.

O'Dea, R. E., et al. 2021. Towards open, reliable, and transparent ecology and evolutionary biology. BMC Biology 19:68.

Poisot, T., C. Guéveneux-Julien, M.-J. Fortin, D. Gravel, and P. Legendre. 2017. Hosts, parasites and their interactions respond to different climatic variables. Global Ecology and Biogeography 26:942–951.

Prudic, K. L., K. P. McFarland, J. C. Oliver, R. A. Hutchinson, E. C. Long, J. T. Kerr, and M. Larrivée. 2017. eButterfly: leveraging massive online citizen science for butterfly conservation. Insects 8:53.

Robinson, O. J., V. Ruiz-Gutierrez, M. D. Reynolds, G. H. Golet, M. Strimas-Mackey, and D. Fink. 2020. Integrating citizen science data with expert surveys increases accuracy and spatial extent of species distribution models. Diversity and Distributions 26:976–986.

Sólymos, P., S. M. Matsuoka, E. M. Bayne, S. R. Lele, P. Fontaine, S. G. Cumming, D. Stralberg, F. K. A. Schmiegelow, and S. J. Song. 2013. Calibrating indices of avian density from non-standardized survey data: making the most of a messy situation. Methods in Ecology and Evolution 4:1047–1058.

Soroye, P., et al. 2022. The risks and rewards of community science for threatened species monitoring. Conservation Science and Practice 4:e12788.

Sousa, L. B., A. Craig, U. Chitkara, S. Fricker, C. Webb, C. Williams, and K. Baldock. 2022. Methodological diversity in citizen science mosquito surveillance: a scoping review. Citizen Science: Theory and Practice 7:1–19.

Strimas-Mackey, M., S. Ligocki, T. Auer, and D. Fink. 2021. Ebirdst: tools for loading, plotting, mapping and analysis of eBird status and trends data products. https://cornelllabofornithology.github.io/ebirdst/

Strimas-Mackey, M., E. Miller, and W. Hochachka. 2018. Auk: EBird data extraction and processing with AWK. https://cornelllabofornithology.github.io/auk/

Strydom, T., et al. 2022. Food web reconstruction through phylogenetic transfer of low-rank network representation. Methods in Ecology and Evolution 13. https://doi.org/10.1111/2041-210X.13835

Sullivan, B. L., et al. 2014. The eBird enterprise: an integrated approach to development and application of citizen science. Biological Conservation 169:31–40.

Sutherland, W., et al. 2009. One hundred questions of importance to the conservation of global biological diversity. Conservation Biology 23:557–567.

Troudet, J., P. Grandcolas, A. Blin, R. Vignes-Lebbe, and F. Legendre. 2017. Taxonomic bias in biodiversity data and societal preferences. Scientific Reports 7:9132. https://doi.org/10.1038/s41598-017-09084-6

Wilkinson, M. D., et al. 2016. The FAIR guiding principles for scientific data management and stewardship. Scientific Data 3:160018.

Zipkin, E. F., B. D. Inouye, and S. R. Beissinger. 2019. Innovations in data integration for modeling populations. Ecology 100:1–3.