Review

The eBird enterprise: An integrated approach to development and application of citizen science

CrossMark

Brian L. Sullivan [a],[*], Jocelyn L. Aycrigg [b], Jessie H. Barry [a], Rick E. Bonney [a], Nicholas Bruns [a]
Caren B. Cooper [a], Theo Damoulas [c], André A. Dhondt [a], Tom Dietterich [d], Andrew Farnsworth [a]
Daniel Fink [a], John W. Fitzpatrick [a], Thomas Fredericks [a], Jeff Gerbracht [a], Carla Gomes [c]
Wesley M. Hochachka [a], Marshall J. Iliff [a], Carl Lagoze [e], Frank A. La Sorte [a], Matthew Merrifield [f]
Will Morris [a], Tina B. Phillips [a], Mark Reynolds [f], Amanda D. Rodewald [a], Kenneth V. Rosenberg [a]
Nancy M. Trautmann [a], Andrea Wiggins [g], David W. Winkler [h], Weng-Keen Wong [d]
Christopher L. Wood [a], Jun Yu [d], Steve Kelling [a]

[a] Cornell Lab of Ornithology, 159 Sapsucker Woods Road, Ithaca, NY 14850, USA
[b] National Gap Analysis Program, University of Idaho, 530 S. Asbury Street, Moscow, ID 83843, USA
[c] Department of Computer Science, Cornell University, 5133 Upson Hall, Ithaca, NY 14853, USA
[d] Oregon State University, 1148 Kelley Engineering Center, Corvallis, OR 97331, USA
[e] University Of Michigan School of Information, 115 S. State Street, Ann Arbor, MI 48103, USA
[f] The Nature Conservancy, California Program, 201 Mission Street, 4th Floor, San Francisco, CA 94105, USA
[g] University of New Mexico, 1312 Basehart SE, Albuquerque, NM 87106, USA
[h] Department of Ecology and Evolutionary Biology, Museum of Vertebrates and Lab of Ornithology, Cornell University, Ithaca, NY 14853, USA

ARTICLE INFO

ABSTRACT

Citizen-science projects engage volunteers to gather or process data to address scientific questions. But citizen-science projects vary in their ability to contribute usefully for science, conservation, or public policy. eBird has evolved from a basic citizen-science project into a collective enterprise, taking a novel approach to citizen science by developing cooperative partnerships among experts in a wide range of fields: population and distributions, conservation biologists, quantitative ecologists, statisticians, computer scientists, GIS and informatics specialists, application developers, and data administrators. The goal is to increase data quantity through participant recruitment and engagement, but also to quantify and control for data quality issues such as observer variability, imperfect detection of species, and both spatial and temporal bias in data collection. Advances at the interface among ecology, statistics, and computer science allow us to create new species distribution models that provide accurate estimates across broad spatial and temporal scales with extremely detailed resolution. eBird data are openly available and used by a broad spectrum of students, teachers, scientists, NGOs, government agencies, land managers, and policy makers. Feedback from this broad data use community helps identify development priorities. As a result, eBird has become a major source of biodiversity data, increasing our knowledge of the dynamics of species distributions, and having a direct impact on the conservation of birds and their habitats.

© 2013 Elsevier Ltd. All rights reserved.

Contents

* Corresponding author. Tel.: +1 609 694 3280.
  E-mail address: bls42@cornell.edu (B.L. Sullivan).

1. Introduction

Citizen science is gaining momentum (Irwin, 1995; Dickinson et al., 2010), but citizen-science projects vary in their ability to deliver useful data for research and conservation. Tulloch et al. (2013) reviewed the efficacy of various citizen-science approaches, and recommended a series of actions that could improve the impact and utility of citizen-science data. These included incorporating more rigorous sampling techniques into unstructured data collection processes, improving data quality, broadening the data-user community, and improving communication between those using the data and those collecting it. We suggest that achieving these objectives is best accomplished by expanding the range of activities routinely encompassed with the running of citizen-science projects to extend beyond data collection to include community engagement, data curation, data synthesis and analysis, pattern visualization, and delivery of results to a broad community of possible stakeholders. In this paper we will illustrate this point using eBird (ebird.org; Sullivan et al., 2009) as our example. eBird has evolved from a stand-alone citizen-science project focused on collecting data, into a cooperative partnership involving several distinct user groups spanning multiple scientific domains and dozens of partner organizations, each with its own unique interests in the project's success.

eBird collects information about the distribution and abundance of birds, taking advantage of the enormous popularity of watching birds to create a global network of volunteers who submit bird observations via the Internet to a central data repository. Through a combination of broad-based community engagement and global partnerships, the volume of data coming into eBird has increased exponentially (30–40% annually) for a decade (Fig. 1). By mid-year 2013, over 140 million observations had been submitted by 150,000 separate observers, who spent 10.5 million hours in the field collecting data. While global in scope, the majority of eBird data still come from North America.

Not only has input of data grown, but output of information has grown as well, fostered by collaboration among researchers from diverse scientific domains including ornithology, landscape ecology, macroecology, biogeography, computer science, statistics, computational sustainability, human computation, informal science education, conservation science, and public policy. Engagement in eBird across these domains creates continuous investment and feedback from a diverse community that helps shape eBird to meet the needs of a wide array of both academic and non-academic audiences. This communication is aided by eBird's integrated project structure involving participants who collect data, researchers analyzing the data, agencies adopting policies based on the data, and land managers taking direct
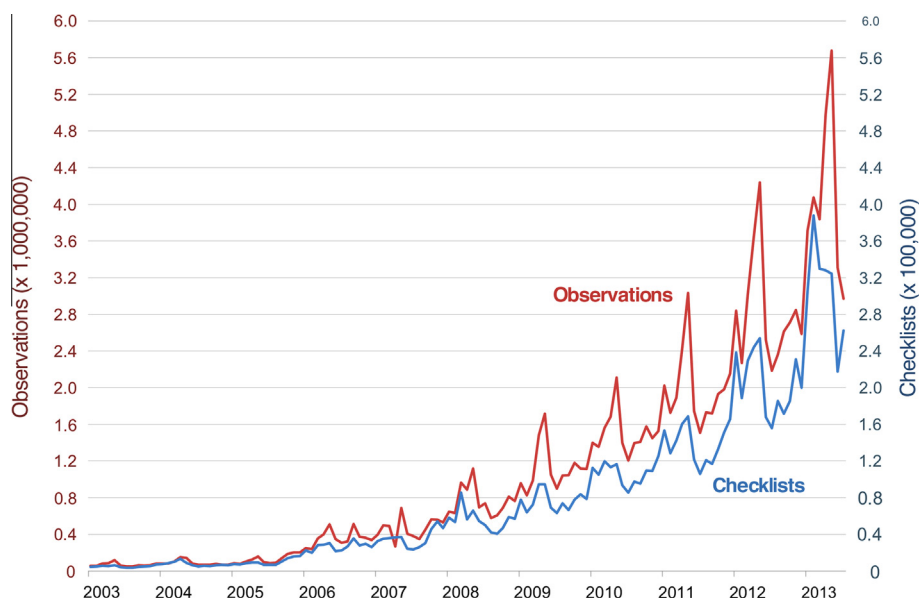

Fig. 1. eBird Growth. The number of observations and checklists submitted monthly since 2003. Note within-year fluctuations, with contributions dropping in summer and increasing through May.
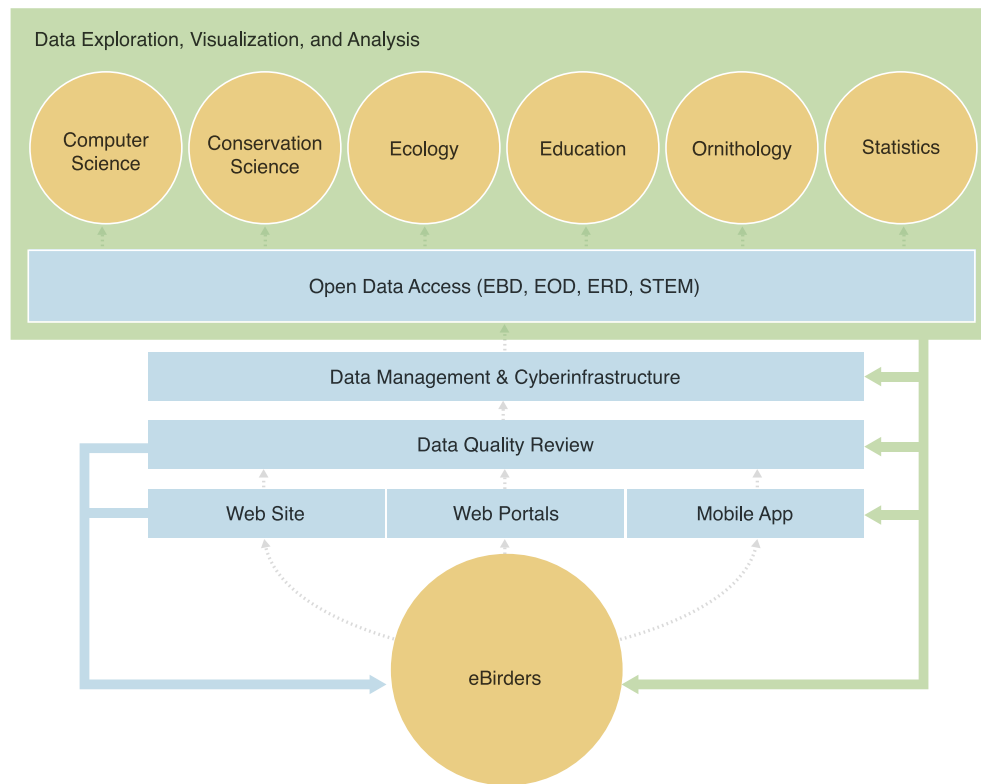
Fig. 2. Data flow, interdisciplinary integration, and feedback in the eBird enterprise. User communities are represented by circles; hardware, software, and infrastructure by rectangles. Starting at the bottom of the figure, eBirders enter observations either through the core eBird web site, an eBird portal, or the mobile application. All data then pass through one centralized data quality process before entering the centralized data management and cyberinfrastructure. The database infrastructure provides the hub for all data storage, curation, and retrieval. Data exploration, visualization, and analysis are provided through close collaborations with the one or more domains. Four data products are created (EOD, EBD, ERD, and STEM), which are openly accessible and tracked, leading to broad scale data use. The feedback loops are key. All aspects of the eBird enterprise inform eBird development and provide feedback to the user in some way (green or blue arrows). For example, our research and analysis teams suggest best practices for data collection, and our data quality process helps participants realize when they've made a mistake, helping them become better birders. Our web tools are geared toward providing user reward. All of these things lead to increased engagement and data volume, which in turn creates better data products and more informed conservation outcomes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

conservation action (Fig. 2). Ultimately, these collaborations have enabled us to increase both the quantity and the quality of useful data available for analysis. Further, this broad spectrum of intellectual contributions and applications has fundamentally changed our view of the project, which we now see as a collective enterprise.

2. Increasing data quantity while improving data quality

Hochachka et al. (2012) showed that maximizing the analytical value of citizen-science data requires striking the right balance between data quantity and data quality. Large quantities of data are needed because the per-datum information content of citizen-science data is often low, because most projects use relatively simple protocols. However, we also need to identify minimal standards of quality, and ideally raise these standards through time. Thus, we have focused on both growing the data-collecting community engaged with eBird, while simultaneously improving the project's data quality mechanisms through a variety of means.

2.1. Meeting community needs

Increasing the number of participants in a project and the amount of information that each participant provides requires identifying this audience and meeting their needs. For eBird's audience of active birdwatchers, the ability to rapidly improve the technologies that we use to interact with participants is key. In

2011, we redesigned eBird's online data entry process to improve efficiency and flexibility. The new process allows users to enter data more quickly, but also features a flexible data collection framework that promotes submission of ancillary data (e.g., breeding status, age and sex classes). At the same time, feedback solicited from participants has shaped new developments that provide valuable end-user tools and rewards. For example, participants can keep track of their personal bird lists, receive rare bird alerts, explore information on when and where to find birds, and take part in games that appeal to competitive spirit. As a result, participants become increasingly invested in the project, thereby increasing data volume (Wood et al., 2011).

We also recognize that a growing number of people are no longer using desktop computers as their primary means of connecting to the Internet. In 2012, we partnered with a smart phone application developer to release "BirdLog" (birdseyebirding.com), an app that allows anyone to submit or retrieve eBird observations on mobile devices anywhere around the world. Currently nearly 20% of eBird data are submitted on mobile devices in the field.

Keeping participants' bird lists updated and accurate is another critical component of eBird's community services. As with any organismal database, detailed attention to taxonomy and nomenclature is essential. Thus, maintaining a current, accurate, and user-friendly taxonomy is an integral part of eBird. eBird maintains its own taxonomy for use by birdwatchers that is updated annually (help.ebird.org/customer/portal/articles/1006825), keeping participants' records current with changes in our knowledge of how many species of birds exist and where they live.

2.2. A flexible platform

Another aspect of meeting participants' needs is making adaptations for specific audiences. eBird's user interface is designed to allow a level of customization to suit a variety of potential use cases. For example, we aim to overcome language barriers to participation by making eBird available through customized "portals", which can be created in any non-English language and managed by partner organizations outside of the United States (e.g., averaves.org) (Wood et al., 2011). Specific interest groups can be engaged, such as the Young Birders Network—a portal managed and written by young people who are passionate about birds (ebird.org/content/ybn). This flexibility ensures a high level of local expertise, promotion, and project ownership to better engage regional audiences.

The data collection process can also be modified to meet specific research goals. Customizing and augmenting eBird's standard data collection protocols promotes a level of adoption that would be impossible if eBird maintained a rigid approach to data collection. Importantly, in spite of different routes for the entry of information, the data coming in through each customized portal are fully integrated with the eBird database and application infrastructure so that information can be shared and analyzed freely across linguistic, political, and geographic boundaries. This integrated application infrastructure ensures that enhancements in one portal are available in all other portals, providing economies of scale and access to database and application infrastructure that would otherwise be too costly for any single institution to implement.

Many users of eBird are not interested in developing partnerships and portals, but instead use eBird as an informal cyberinfrastructure for data ingestion and output supporting their work. In many cases, the individuals and groups coordinating such projects have no resources for technology support (Wiggins, 2013), and adopt the basic functionality of eBird to fulfill this function. Others have recognized that eBird provided substantial efficiencies for tasks that would otherwise require considerable additional coordination and development overhead (Wiggins, 2012).

2.3. Improving data quality

While the straightforward data collection protocols and participation incentives employed by eBird engage large numbers of participants, these protocols also produce data with higher levels of noise than data collected under more rigorous protocols. Several important sources of bias have been discussed in the literature including variation among observers (Cooper et al., 2007; Yu et al., 2010), "false absences" owing to imperfect detection of organisms (McClintock et al., 2010), and uneven distributions of data in space and time (Boakes et al., 2010; Chu and Dietterich, 2013). Some of this variation can be accounted for during data analysis, but other types of noise need to be identified and removed as part of the data assimilation process. In this section we review steps we have taken to improve the quality of data collected by eBird.

eBird's process for maintaining data quality is a developing example of human computation (the engagement of human intelligence in targeted, programmatic tasks (Law and Ahn, 2011)), which seeks to identify and use the complementary strengths of machine and human intelligence. We identify eBird's combination of human and machine intelligence as a Human/Computer Learning Network (Kelling et al., 2013), with feedback between humans and computers allowing continued improvements in quality control. eBird uses a two-part approach to quality control during data entry: automated filters and a growing network of regional editors. Automated data quality filters flag records for review based on observation date and geographic location. This process provides the user with immediate feedback during checklist submission, allowing the user to correct typographical errors or reconsider a given species' identification. Once a flagged entry is confirmed as legitimate by the observer, a regional expert reviews the record, communicating with the observer to acquire and vet supporting documentation. Reviewers provide active feedback to participants on everything from bird identification to best practices for eBird data collection. This relationship between reviewers and participants leads to increased participant skill, thereby generating more useful data for analysis (Sullivan et al., 2009). The spatial and temporal resolution of the automated filters has become continually finer since eBird's inception, and eBird is now moving toward a data-driven approach to creating filters. Using existing data to identify threshold dates and counts that would be considered unusual for each species, we are creating more accurate spatial and temporal filters for vetting incoming checklist submissions. We further improve these filters by providing a tool that allows expert regional editors to fine-tune the frequency thresholds generated by this "emergent filter" process (Kelling, 2011). This combined process has reduced the workload for volunteer expert reviewers by decreasing the number of inappropriately flagged records, and improves the end-user experience during data submission by providing more accurate feedback on what constitutes an unexpected species or unusually high count.

3. Access to eBird data

Successful citizen science does not end at data collection; data management and delivery are also critical. eBird data are maintained in a highly managed centralized data infrastructure that serves as a benchmark among citizen-science programs (Kelling, 2011). A core principle of eBird is providing free and open access to data for a broad spectrum of data consumers. This diversity of users and uses of eBird data requires that we develop and curate multiple eBird data products to meet the specific needs of varied audiences. Two of these data products are based on the core data fields within the eBird database. Two additional eBird data products provide access to eBird data integrated with supplemental environmental data geared toward analysis and modeling.

3.1. eBird data access policy

A growing number of biodiversity databases and database clearinghouses provide access to data. However, proper attribution and acknowledgment for data providers is often lacking both by the clearinghouses as well as end-users. This issue led us to institute a new data access policy for eBird. Beginning in November 2012, we required anyone interested in downloading certain eBird data products to register with eBird. While eBird data are free, we require registration in order (1) to encourage communication between potential users and eBird staff, (2) to provide the Cornell Lab of Ornithology with metrics that allow analysis of data usage trends, (3) to improve data products, and (4) to create a method by which users can attest that eBird data will not be misused or added to commercial products. Data users must agree to the eBird terms of use policy, provide their name, country of residence, identify their project type, provide the title of their project and an abstract, as well as their affiliation. These requests are individually reviewed to ensure validity, after which a data access key is granted. Once registered, users can access a series of online tools that allow queries to be written against the database to extract target datasets. eBird data are bundled with our terms of use policy, metadata descriptors, and recommended citations.

3.2. Core eBird data products

The eBird observational dataset (EOD) contains primary species occurrence data defined as a record of a particular taxon in a particular place at a particular point in time (Soberon and Peterson, 2009). eBird data consisting of location, date, and species are provided in the EOD to ensure that they can be integrated with other observational data and natural history collections data for estimating patterns in biodiversity (Jetz et al., 2012). The EOD is updated annually and is freely accessible via DataONE (dataone.org), the Global Biodiversity Information Facility (gbif.org), the Avian Knowledge Network (avianknowledge.net), and VertNet (vertnet.org). The 2012 version contains more than 100 million bird observation records. Although data usage reports of the EOD are not provided by these data clearinghouses, EOD data from GBIF are used in other initiatives such as the NSF funded Lifemapper Project (lifemapper.org), and the United States Geological Survey's Biodiversity Information Serving Our Nation (BISON) facility (bison.usgs.ornl.gov), and use of the EOD in ecological studies is widespread (e.g., Klicka et al., 2011; Lait et al., 2012; Davis et al., 2013).

The eBird basic dataset (EBD) contains data in the form of checklists: counts of all bird species observed during a single search event (Sullivan et al., 2009). The EBD contains counts of species detected, the location where the search took place, the time the search was initiated, and the duration and length of the search. One value of grouping species within sets checklists in the EBD is the ability to identify associations between species. The EBD is updated quarterly and is available on the eBird website. As of August 2013 the EBD contained 134 million observation records.

3.3. Value-added eBird data products

Many ecological and conservation problems require comprehensive, high-resolution information about the spatial and temporal distribution of populations. This poses a challenge because species observation data are often sparsely and irregularly distributed, especially for widely distributed species. To address this challenge, species distribution models are used to estimate species' distributions in unobserved locations and times. In this section we describe two "value-added" eBird data products used for species distribution modeling and analysis.

The eBird reference dataset (ERD) consists of the EBD with two additional types of information: species absence information and a suite of variables that describe the local environment. Species absence information is included to facilitate the control of biases associated with the imperfect detection of birds. The importance of controlling for biases in detection with wildlife survey data is well known (MacKenzie et al., 2006), and is typically dependent on knowing when a species was observed as well as when it *was not* observed. By limiting the ERD to the set of checklists where participants confirmed that they were reporting all species identified, zero counts can be included for unreported species. These zeros are sometimes referred to as "pseudo-absences" and indicate a lack of detection or identification. The pseudo-absences are valuable because they convey partial information about absence, which can be combined with participant-recorded information on search effort to capture and control for sources of variation associated with the detection process. The second set of variables included in the ERD describes the local environment where searches took place (e.g., landcover and climate). This information is provided to link species' distributions to the environments on which they depend. The ERD is updated annually, covers the western hemisphere, and contains 89 million observations. The ERD is available via the eBird web site.

By relating environmental predictors to observed occurrences, species distribution models can make predictions at unsampled locations and times. Elith and Leathwick (2009) provide a good review of the literature on species distribution models, using conventional analytical methods. Our early work with such methods revealed that they were not able to consistently produce accurate distribution models, and the creation of novel analytical techniques (Spatio-Temporal Exploratory Models (STEM); Fink et al., 2010, 2013) has led to an extensive engagement of statisticians and computer scientists in the eBird project. The resultant STEM models currently provide seasonal estimates of species' distributions across the contiguous U.S. for more than 300 species. For each species the probability of occurrence is estimated for each week of the year based on local landcover characteristics at approximately 1 million locations selected from a stratified random design.

In addition, the STEM distribution occurrence estimates are "adjusted" to control for variation in detectability associated with search effort. By relating predictors that describe the detection process (e.g., search effort) to observed occurrences, species distribution models can control for these confounding sources of variation (e.g., Link and Sauer, 1999; Lele et al., 2012). To facilitate this analysis, the ERD dataset only includes checklists where users reported effort variables: search time, transect length, time of day, number of observers, and protocol type. These variables are used to adjust the distribution estimates so that they represent a survey with constant effort; specifically, by estimating the probability that a typical eBird participant will detect the species on a search from 7 to 8AM while traveling 1 km on the given day at the given location.

The model-based distributional estimates from STEM provide a good source of population-level distributional data for further studies. Visualizations for many species are available on the eBird website (ebird.org/plone/ebird/about/occurrence-maps). The STEM model distributional data are available through collaboration and partnership with the Cornell Lab of Ornithology. In the future, the geographic scope of these distributional models will expand beyond the contiguous U.S.

4. Diverse uses of eBird data

By combining the strengths and motivation of volunteer observers with computer technologies, the eBird project is generating biodiversity data for an increasing variety of uses. In the past decade, more than 90 peer-reviewed publications either used eBird data or studied aspects of the eBird project. Over the past year, more than 1100 individuals from 40 countries have requested eBird data, making more than 3400 downloads representing ca. 2.6 terabytes of data. Data users were classified into four broad categories (Fig. 3a). Almost 50% of requests came from individuals, self-categorized as academic or student, who were looking for datasets for research projects. While most of these students were studying species distribution, a significant number were using the eBird data to support projects in GIS, statistics, or computer science (Fig. 3b). Many NGO/governmental users were using eBird data for estimating species occurrences on both public and private landholdings. Members of the general public mostly use eBird data to explore bird occurrences in their region of interest (i.e., general use), while commercial users were doing environmental impact statements for local or regional governments.

Such broad data use is a major part of eBird's success, and it inspires participation in the project, both from the birding and research communities. In this section we discuss how data are being used within four key communities, and how collaboration and communications with these groups are further improving both collection and use of the data.
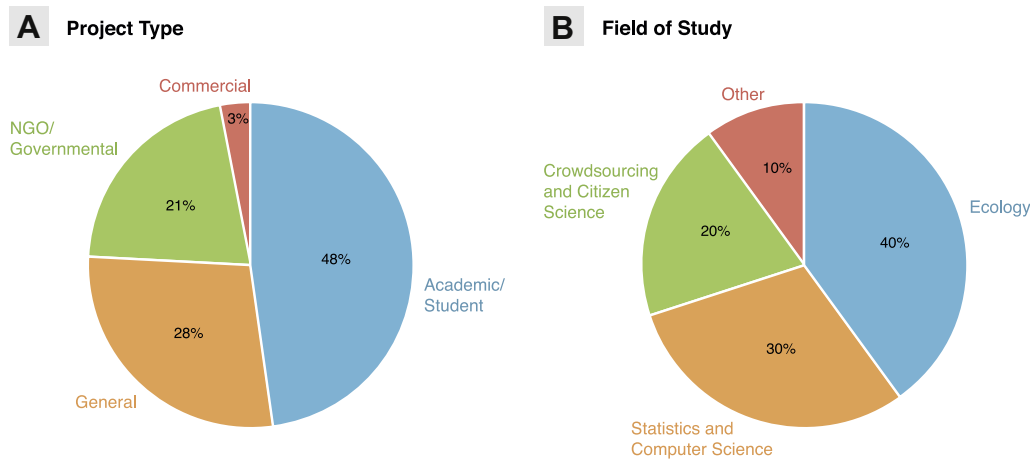
A Project Type

B Field of Study

Fig. 3. eBird data use based on 1100 data requests. Understanding the data user community is key in developing an effective citizen-science project, and delivering useful data for research and conservation. Feedback from these users helps shape development of eBird data products.

4.1. Advancing computational methods

Key to the eBird enterprise is the diverse research group that is actively forming around it. Consisting of scientists from a wide array of backgrounds, the research conducted by this group has improved eBird in numerous ways, ranging from the development of robust data products, to techniques for detecting and accounting for sampling bias, to helping eBird's project leaders inform participants on data entry best practices, and how to increase the research value of the data.

The increasing need for sophisticated statistical models and computational techniques that integrate, interpret, and analyze these vast data resources has led the eBird research group to pursue several data-intensive research challenges. Unique challenges associated with analyzing eBird data have led to innovations in machine-learning and statistics (Sorokina et al., 2007; Sheldon, 2010; Fink et al., 2010, 2013; Hutchinson et al., 2011). Novel approaches to species distribution modeling (Fink et al., 2010, 2013) resulted in the creation of two of eBird's data products: the eBird Reference Data Set, and the eBird Spatio-Temporal Exploratory Models. In addition, research extending the standard "Occupancy and Detection" species distribution model (MacKenzie et al., 2002) has improved our understanding of participant observational skills (Hutchinson et al., 2011; Yu et al., 2010). This, in turn, will improve accuracy and analytical power in all fields where citizen-science or expertise-based information is used, especially in ecology, eco-informatics, and computational sustainability.

4.2. Increasing understanding of avian migration ecology

eBird provides unique opportunities to investigate questions in western hemisphere avian ecology with unprecedented spatial and temporal detail (Hochachka and Fink, 2012). One area to which eBird is particularly well suited is the study of migration ecology for medium- and long-distance migrants. Typically, only the movements of individual birds in migration have been studied using band returns or, more recently, tracking devices (Bairlein, 2003). eBird provides the opportunity to study avian migration from a population-level perspective with a comprehensive spatial and temporal representation of migration of entire populations rather than individuals. For example, we can now quantify migration patterns and speeds at the population-level for many species across their annual cycles (Fig. 4). This has allowed for the testing of key predictions originating from optimal migration theory (La Sorte et al., 2013) and, for the first time, has allowed us to quantify

the structure, seasonal dynamics, and determinants of migration flyways (La Sorte et al., Submitted for publication). In total, eBird is providing opportunities for major advances in how we understand and study the ecology and evolution of avian migration. As many populations of long-distance migratory species are declining (Robbins et al., 1989; Sanderson et al., 2006), this work also supports modeling and mitigation efforts that are addressing the long-term implications of global climate change for migratory birds (La Sorte and Jetz, 2010).

4.3. Informing conservation policy and action

The North American Bird Conservation Initiative (NABCI) partnership has produced a series of comprehensive analyses of the state of the birds for the United States (stateofthebirds.org). Because of the recognition that bird populations are useful indicators of ecological health and biodiversity status (Pereira and Cooper, 2006; Lawton et al., 1998; Butchart et al., 2010), these reports have broad significance for national conservation planning. The two recent State of the Birds reports used USGS-GAP's Protected Areas Database of the U.S. (PAD-US) and eBird (i.e., EOD and STEM data) to estimate the stewardship responsibilities for bird populations on both public and private lands (U.S. NABCI, 2011, 2013).

A long-standing challenge in bird conservation has been the difficulty of studying complete annual cycles for any bird species that undergoes broad-scale and complex geographic movements. The rich spatial and temporal detail of eBird is beginning to break this barrier, facilitating analyses that provide information across large components of species' annual cycles. For example, the use of eBird STEM with the US Protected Areas Database provided the opportunity to estimate weekly stewardship responsibilities for 370 species on public and private lands within the contiguous U.S. By combining these data, La Sorte et al. (Submitted for publication) identified species that moved seasonally between lands under different public ownership, which provides a compelling rationale for inter-agency cooperation to develop full life-cycle conservation plans for birds.

In partnership with The Nature Conservancy of California, we are currently summarizing information needed to identify seasonally varying priority conservation areas for target species of migratory birds, in order to design effective conservation strategies. California's Central Valley is one of the most altered landscapes in the world, with only 5% of the former extent of 4 million acres of wetlands remaining (Hoekstra et al., 2005). Nevertheless the refuges, managed wetlands, and compatible agricultural lands in the
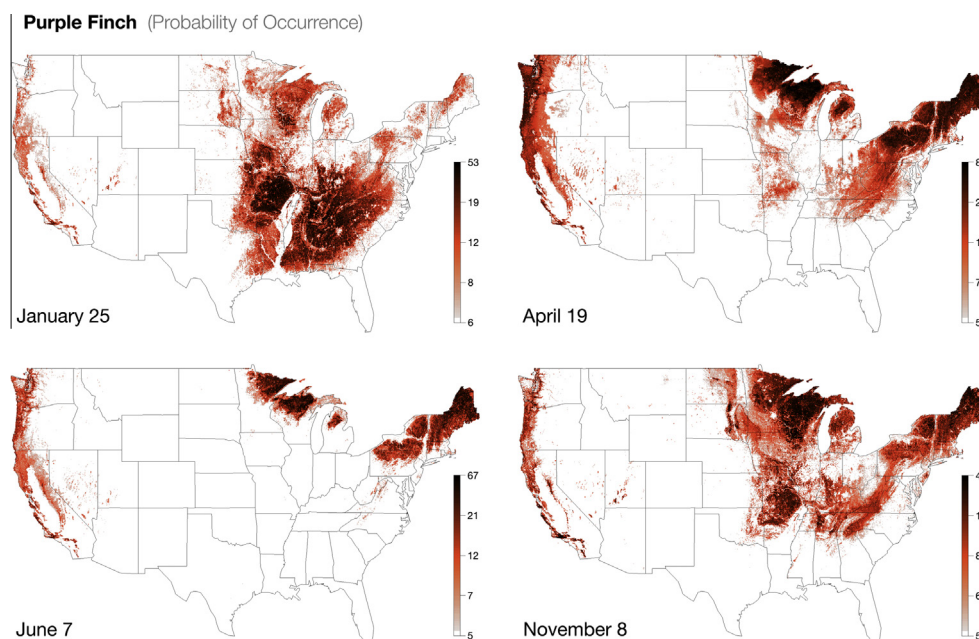
Purple Finch (Probability of Occurrence)

Fig. 4. STEM species distribution model for Purple Finch (*Haemorhous purpureus*) based on ERD data. Darker red indicates higher probability of finding a Purple Finch (Max = 80%). Note population-level differences in migration, with the eastern/boreal population (*H. p. purpureus*) migrating and wintering through the East and Southeast, whereas the western population (*H. p. californicus*) is largely resident, showing little seasonal change in distribution. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

region still support millions of migrating and wintering waterfowl and shorebirds of the Pacific Flyway (Western Hemisphere Shorebird Reserve Network, 2003; Central Valley Joint Venture, 2006). In a region prone to drought, changing land uses, and water allocation disputes, specific and actionable recommendations for annual and long-term habitat management and water allocation are critical. The Nature Conservancy's goal is to use STEM models to identify high value areas for birds, and then work with landowners and managers to better manage these key sites for water birds. Already, this work has led to the creation of effective conservation easements, habitat management programs, and metrics for habitat credit trading and mitigation programs.

4.4. Education: student investigations with eBird

Science education initiatives, from grade school through undergraduate levels, call for students to pose scientific questions, design investigations, analyze data, and interpret the results. eBird fits well with this mandate, and students and educators use eBird in classroom and independent projects ranging from ornithology to statistics (Fee and Trautmann, 2012; Trautmann et al., 2012; Trautmann and McLinn, 2012). Accompanying curriculum resources have been created to support teachers in facilitating student investigations (Schaus et al., 2007; Fee et al., 2011, 2013). Before or after observing local birds, students can use eBird data to determine which species are common in their community, discover trends, and develop hypotheses. Whether using their own bird observations or looking for patterns in online data collected by others, students can pose research questions and conduct investigations addressing questions of their choice. In this era of big data, eBird gives students the opportunity to envision research beyond stereotypical images of individual scientists working in the lab or field. Using the animated "occurrence maps" that model migration dynamics, they see visually striking evidence for the relevance of citizen science in determining species-specific habitat requirements throughout the year (Fee, 2013). Such use of online data visualization tools is a natural fit for today's students, most of

whom have grown up using a wide range of digital applications in their daily lives and will rely on them increasingly throughout their careers.

5. The future

We anticipate a series of key future developments that will help all facets of the eBird enterprise continue to evolve.

5.1. Making global connections—Tringa   → Conclusão

We envision a global bird monitoring network that engages the interests of birders across countries and cultures. To achieve this requires significant investment in international partnership-building, leveraging the knowledge and relationships of partners to help develop critical community engagement. While eBird's tools are available worldwide, adoption in many regions has been marginal. In some cases, other citizen-engagement projects are already in place, with local institutions and birders already doing an excellent job collecting data. In many cases, however, these systems are idiosyncratic and regionalized. Some only track information on resident birds, or on migratory species during certain times of the year. The data from all of these local observational data systems could, in theory, be brought together under a common framework so that data can be shared and used for conservation efforts at global scales. To this end, a new initiative is underway called "Tringa" (named after a genus of shorebird that occurs worldwide), intended to leverage the data collecting and outreach capabilities of organizations around the world by making all existing data available within a common data framework. The concept is currently being tested through collaboration between Cornell Lab of Ornithology with its eBird project, and the British Trust for Ornithology who runs a comparable project in the United Kingdom called BirdTrack (blx1.bto.org/birdtrack/main/data-home.jsp). Our hope is that as Tringa evolves, more partners will join from around the world, expanding the coverage of the shared database.

5.2. Community engagement through tool development

eBird's tools thus far have been developed to serve a largely North American audience, but with increasing use around the world, we must shift and adapt our development to engage communities that have different interests and goals. For example, the advent of inexpensive, high-resolution digital cameras has enabled a new generation of nature enthusiasts to experience birds in ways never before possible. We believe that eBird must build the partnerships required to engage these existing communities, as well as develop its own tools to allow participants to store and share rich media. The Macaulay Library of Natural Sounds (macaulaylibrary.org/) has a long history of careful curation of rich media, and serves as a permanent archive for millions of sounds, videos, and photos. By leveraging new technology and the curatorial capacity and standards of the Macaulay Library, eBird will soon be a massive conduit for the collection and sharing of rich media.

eBird relies on participants to identify birds correctly, but for most people, identifying birds is at times challenging. Thousands of species must be considered, and most reference materials require knowledge of specific terminology in order to be helpful. "Merlin" (beta URL: allaboutbirds.org/labs) is an online bird identification tool that is being developed to lower the barrier to bird identification for a broad audience. Merlin asks the user a series of simple questions including the location, date, color, and size of an unknown bird. By combining date- and place-specific eBird data and crowd-sourced compilations of information on how people describe birds, with machine learning analytical techniques, Merlin computes a list of likely matches and presents the list to the user. eBird data serves as the backbone for Merlin, enabling it to calculate the probability of species occurrence at locations across North America, each day of the year.

5.3. Novel insights from analysis

Effective conservation requires understanding the needs of birds throughout their full annual cycles, which is particularly challenging for species that require separate breeding and non-breeding grounds and are dependent on migratory stopover locations (Hagan and Johnston, 1992). Migration and winter represent a large proportion of many species' lives, yet we still know remarkably little about birds' requirements during these critical phases of the annual cycle. eBird data and models are showing when and where species occur throughout the western hemisphere, revealing population-level movements and habitat associations of birds across their full annual cycles (Fig. 5). Importantly, these models will also reveal how habitat use changes through space and time.

A current focus of ours is development of analytical methods for accurately estimating species abundance across space and time from eBird data, not simply modeling probabilities of occurrence. As with our work to model bird occurrences, we need to evaluate, adapt, and extend existing approaches for abundance modeling (e.g., Royle and Dorazio, 2008; Sauer and Link, 2011) in order to deal with the issue of non-stationarity of associations through space and time, and to ensure that analyses can be run with limited human intervention. We can assume that, in general, a species that is more abundant is more likely to be seen and reported, but the nature, variability, and species-specific nuances of that relationship are unclear. For species that occur in large aggregations, describing fine-scale variation in abundance is critical for effective conservation. Even less dramatic variation in abundance can go undetected in analyses of data describing occurrence. For example, if a resident species' abundance at a particular location increases, perhaps as the result of an irruption (e.g., Black-capped Chickadee (*Poecile atricapillus*)), or decreases as a result of an emerging
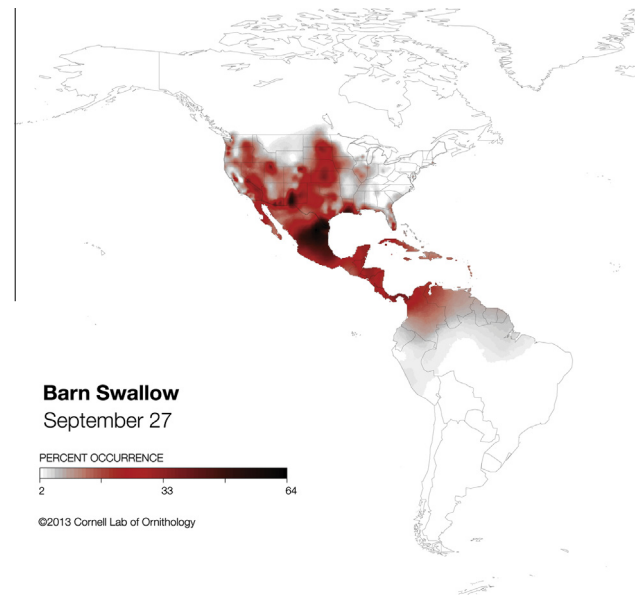

Fig. 5. Preliminary AdaSTEM species distribution model for Barn Swallow (*Hirundo rustica*) during fall migration across the western hemisphere. eBird's broad spatial and temporal scale allow modeling species across the complete life cycle. Importantly, these models also analyze species-habitat relationships that change through space and time.

disease (e.g., *Mycoplasma gallisepticum* in House Finches (*Haemorhous mexicanus*) (Hochachka and Dhondt, 2000)), current analyses of eBird data will not effectively show this. Estimating abundance across large parts of the globe, throughout the year, will allow for increased application of eBird data for research and conservation.

Knowledge of migratory behavior is critical for studying the ecology and evolution of bird movements (Greenberg and Marra, 2005; Salewski and Bruderer, 2007). This information can be used in applying best conservation practices for migratory birds. A new research project called BirdCast (birdcast.info) is unifying eBird and weather surveillance radar data to predict the species composition, timing, magnitude, speed, and direction of bird movements across the continental US. By leveraging innovative machine-learning techniques to develop models to identify complex conditions governing the dynamics of migration behavior, choice of migratory pathways, factors that influence when birds migrate, and the speed and duration of nocturnal movements, BirdCast has the potential to transform our understanding of population-level bird movements at the continental scale (Sheldon et al., 2013).

Global climate change, land conversion, globalization, and human population growth are presenting significant conservation challenges that require understanding processes that occur at continental and global scales. Furthermore, the dynamic nature of these processes requires the ability to monitor systems across full annual cycles in near real-time to identify current and impending threats, as well as to evaluate the outcomes of conservation actions. However, scientific and conservation communities remain limited by the challenge of collecting the data required to simultaneously support inference and action at multiple spatial and temporal scales. For terrestrial ecological communities at least, eBird is well poised to provide a solution. Some of the most compelling information on the implications of global change for biodiversity has come from the analysis of citizen-science data. For example, data compiled on birds (Hurlbert and Liang, 2012; La Sorte and Jetz, 2012; Monahan and Tingley, 2012) and butterflies (Breed et al., 2013) by citizen scientists in North America have been used to identify the biological consequences of climate change. Management agencies and conservation organizations are beginning to

recognize that eBird provides invaluable year-round data on species distributions and abundance, at range-wide scales, but also with fine spatial and temporal resolution. eBird's importance for providing these data as a component of full life-cycle conservation planning will grow. Expanding the data-delivery capacity of the eBird enterprise to meet these demands will be our primary focus over the next decade.

6. Conclusion

Citizen science has the potential to transform our understanding of many ecological processes. For studies of avian ecology, data from citizen-science projects offer the ability to test hypotheses at spatial and temporal scales never before possible. Due to their popularity, birds lend themselves to this approach perhaps better than any other organism. However, in order to be successful, careful attention must be paid to all aspects of the project, from data collection to data delivery and use. In this paper we describe three key aspects for developing a successful citizen science project: striking a balance between data quantity and quality, facilitating wide use of the data by developing and delivering data products, and engaging a diverse array of collaborators in all aspects of the project. We have illustrated these points by describing the eBird enterprise, which over its first decade has demonstrated success in gathering vast amounts of avian biodiversity data that are now being used across many scientific disciplines.

Acknowledgments

References

Bairlein, F., 2003. The study of bird migrations – some future perspectives. Bird Study 50, 243–253.
Boakes, E.H., McGowan, P.J.K., Fuller, R.A., Chang-qing, D., Clark, N.E., O'Connor, K., Mace, G.M., 2010. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. PLoS Biology 8, e1000385.
Breed, G.A., Stichter, S., Crone, E.E., 2013. Climate-driven changes in northeastern US butterfly communities. Nature Climate Change 3, 142–145.
Butchart, S.H.M., Walpole, M., Collen, B., van Strien, A., Scharlemann, J.P.W., Almond, R.E.A., Baillie, J.E.M., Bomhard, B., Brown, C., Bruno, J., Carpenter, K.E., Carr, G.M., Chanson, J., Chenery, A.M., Csirke, J., Davidson, N.C., Dentener, F., Foster, M., Galli, A., Galloway, J.N., Genovesi, P., Gregory, R.D., Hockings, M., Kapos, V., Lamarque, J.F., Leverington, F., Loh, J., McGeoch, M.A., McRae, L., Minasyan, A., Morcillo, M.H., Oldfield, T.E.E., Pauly, D., Quader, S., Revenga, C., Sauer, J.R., Skolnik, B., Spear, D., Stanwell-Smith, D., Stuart, S.N., Symes, A., Tierney, M., Tyrrell, T.D., Vié, J.C., Watson, R., 2010. Global biodiversity: indicators of recent declines. Science 328, 1164–1168.
Central Valley Joint Venture, 2006. Central Valley Joint Venture Implementation Plan – Conserving Bird Habitat. U.S. Fish and Wildlife Service, Sacramento, CA.
Chu, S., Dietterich, T., 2013. A comparative study of methods for density ratio estimation. Technical Report. ScholarsArchive, Oregon State University.
Cooper, C.B., Dickinson, J., Phillips, T., Bonney, R., 2007. Citizen science as a tool for conservation in residential ecosystems. Ecology and Society 12, 11.

Davis, A.Y., Malas, N., Minor, E.S., 2013. Substitutable habitats? The biophysical and anthropogenic drivers of an exotic bird's distribution. Biological Invasions 1–13. http://dx.doi.org/10.1007/s10530-013-0530-z.
Dickinson, J.L., Zuckerberg, B., Bonter, D.N., 2010. Citizen science as an ecological research tool: challenges and benefits. Annual Review of Ecology, Evolution, and Systematics 41, 149–172.
Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics 40, 677–697.
Fee, J., 2013. Animated maps for animated discussions. Lesson 5, pp. 93–100. In: Trautmann, N.M., Fee, J., Tomasek, T.M., Bergey, N.R. (Eds.), Citizen Science: 25 Lessons that Bring Biology to Life, 6–12. NSTA Press, Arlington, VA, pp. 6–12.
Fee, J.M., Rosenberg, A.J., DeRado, L., Trautmann, N.M., 2011. BirdSleuth: Most wanted birds, Version 2. Cornell Lab of Ornithology, Ithaca, NY.
Fee, J., Curley, L., Trautmann, N.M., 2013. Connecting with students through birds. Chapter 4, pp. 23–32. In: Trautmann, N.M., Fee, J., Tomasek, T.M., Bergey, N.R. (Eds.), Citizen Science: 25 Lessons that Bring Biology to Life. NSTA Press, Arlington, VA, pp. 6–12.
Fee, J., Trautmann, N.M., 2012. Connecting to your community through birds and citizen science. Science Scope 36, 62–68.
Fink, D., Hochachka, W.M., Zuckerberg, B., Winkler, D.W., Shaby, B., Munson, M.A., Hooker, G.J., Riedewald, M., Sheldon, D., Kelling, S., 2010. Spatiotemporal exploratory models for broad-scale survey data. Ecological Applications 20, 2131–2147.
Fink, D., Damoulas, T., Dave, J., 2013. Adaptive Spatio-Temporal Exploratory Models: Hemisphere-wide species distributions from massively crowdsourced eBird data. In: Twenty-Seventh AAAI Conference on Artificial Intelligence (AAAI-13). July 14–18, 2013 in Bellevue, Washington, USA.
Greenberg, R, Marra, P.P. (Eds.), 2005. Birds of Two Worlds: The Ecology and Evolution of Migration. JHU Press.
Hagan, J.M. III, Johnston, D.W., 1992. Ecology and Conservation of Migrant Landbirds. Smithsonian Institution Press, 609pp.
Hochachka, W.M., Dhondt, A.A., 2000. Density-dependent decline of host abundance resulting from a new infectious disease. Proceedings of the National Academy of Sciences of the United States of America 97, 5303–5306.
Hochachka, W.M., Fink, D., Hutchinson, R., Sheldon, D., Wong, W.-K., Kelling, S., 2012. Data-intensive science applied to broad-scale citizen science. Trends in Ecology and Evolution 27, 130–137.
Hochachka, W., Fink, D., 2012. Broad-scale citizen science data from checklists: prospects and challenges for macroecology. Frontiers of Biogeography 4, fb_15350.
Hoekstra, J.M., Boucher, T.M., Ricketts, T.H., Roberts, C., 2005. Confronting a biome crisis: global disparities of habitat loss and protection. Ecology Letters 8, 23–29.
Hurlbert, A.H., Liang, Z., 2012. Spatiotemporal variation in avian migration phenology: citizen science reveals effects of climate change. PLoS ONE 7, e31662.
Hutchinson, R., Liu, L.-P., Dietterich, T., 2011. Incorporating boosted regression trees into ecological latent variable models. Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, 1343–1348.
Irwin, A., 1995. Citizen Science: A Study of People, Expertise, and Sustainable Development. Routledge, London, UK, 198 pp.
Jetz, W., McPherson, J.M., Guralnick, R.P., 2012. Integrating biodiversity distribution knowledge: toward a global map of life. Trends in Ecology and Evolution 27, 151–159.
Kelling, S., 2011. Using Bioinformatics in Citizen Science. In: J. Dickinson, R. Bonney (Eds.), Citizen Science: Public Collaboration in Environmental Research. Cornell University Press.
Kelling, S., Lagoze, C., Wong, W.K., Yu, J., Damoulas, T., Gerbracht, J., Fink, D., Gomes, C., 2013. EBird: a human/computer learning network to improve biodiversity conservation and research. AI Magazine 34, 10–20.
Klicka, J., Spellman, G.M., Winker, K., Chua, V., Smith, B.T., 2011. A phylogeographic and population genetic analysis of a widespread, sedentary North American bird: the hairy woodpecker (Picoides villosus). The Auk 128, 346–362.
Lait, L.A., Friesen, V.L., Gaston, A.J., Burg, T.M., 2012. The post-Pleistocene population genetic structure of a western North American passerine: the chestnut-backed chickadee Poecile rufescens. Journal of Avian Biology 43, 541–552.
La Sorte, F.A., et al., Atmospheric conditions and the seasonal dynamics of North American migration flyways, Submitted for publication.
La Sorte, F.A., et al., Stewardship responsibilities on public lands and full lifecycle conservation of North American birds, Submitted for publication.
La Sorte, F.A., Jetz, W., 2010. Avian distributions under climate change: towards improved projections. Journal of Experimental Biology 213, 862–869.
La Sorte, F.A., Jetz, W., 2012. Tracking of climatic niche boundaries under recent climate change. Journal of Animal Ecology 81, 914–925.
La Sorte, F.A., Fink, D., Hochachka, W.M., DeLong, J.P., Kelling, S., 2013. Population-level scaling of avian migration speed with body size and migration distance for powered fliers. Ecology 94, 1839–1847.
Law, J., Ahn, L.V., 2011. Human computation. Synthesis Lectures on Artificial Intelligence and Machine Learning 5, 1–121, Morgan & Claypool Publishers.
Lawton, J.H., Bignell, D.E., Bolton, B., Bloemers, G.F., Eggleton, P., Hammond, P.M., Hodda, M., Holt, R.D., Larsen, T.B., Mawdsley, N.A., Stork, N.E., Srivastava, D.S., Watt, A.D., 1998. Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. Nature 391, 72–76.
Lele, S.R., Moreno, M., Bayne, E., 2012. Dealing with detection error in site occupancy surveys: what can we do with a single survey? Journal of Plant Ecology 5, 22–31.

Link, W.A., Sauer, J.R., 1999. Controlling for varying effort in count surveys: an analysis of Christmas bird count data. Journal of Agricultural, Biological, and Environmental Statistics 4, 116–125.

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

MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L., Hines, J.E., 2006. Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence. Elsevier, Amsterdam.

McClintock, B.T., Bailey, L.L., Simons, T.R., 2010. Unmodeled observation error induces bias when inferring patterns and dynamics of species occurrence via aural detections. Ecology 91, 2446–2454.

Monahan, W.B., Tingley, M.W., 2012. Niche tracking and rapid establishment of distributional equilibrium in the house sparrow show potential responsiveness of species to climate change. PLoS ONE 7.

North American Bird Conservation Initiative, U.S. 2011. The State of the Birds 2011 Report on Public Lands and Waters of the United States of America. Washington, DC.

North American Bird Conservation Initiative, U.S. 2013. The State of the Birds 2013 Report on Private Lands of the United States of America. Washington, DC.

Pereira, H.M., Cooper, D.H., 2006. Towards the global monitoring of biodiversity change. Trends in Ecology and Evolution 21, 123–129.

Robbins, C.S., Sauer, J.R., Greenberg, R.S., Droege, S., 1989. Population declines in North-American birds that migrate to the neotropics. Proceedings of the National Academy of Sciences of the United States of America 86, 7658–7662.

Royle, J.A., Dorazio, R.M., 2008. Hierarchical Modeling and Inference in Ecology: The Analysis of Data from Populations, Metapopulations and Communities. Academic Press, San Diego, California.

Sanderson, F.J., Donald, P.F., Pain, D.J., Burfield, I.J., van Bommel, F.P.J., 2006. Long-term population declines in Afro-Paleartic migrant birds. Biological Conservation 131, 93–105.

Salewski, V., Bruderer, B., 2007. The evolution of bird migration—a synthesis. Naturwissenschaften 94, 268–279.

Sauer, J.R., Link, W.A., 2011. Analysis of the North American Breeding Bird Survey using hierarchical models. Auk 128, 87–98.

Schaus, J. M., Bonney, R., Rosenberg, A.J., Phillips, C.B. 2007. BirdSleuth: Investigating Evidence. Ithaca, NY: Cornell Lab of Ornithology. <http://www.birds.cornell.edu/birdsleuth/investigating-evidence>.

Sheldon, D., 2010. Manipulation of PageRank and Collective Hidden Markov Models, in Department of Computer Science. Cornell University.

Sheldon, D., Farnsworth, A., Irvine, J., Van Doren, B., Webb, K., Deitterich, T., Kelling, S. 2013. Approximate Bayesian Inference for Reconstructing Velocities of Migrating Birds from Weather Radar. AAAI Conference on Artificial Intelligence, North America, June 2013. <http://www.aaai.org/ocs/index.php/AAAI/AAAI13/paper/view/6468>.

Soberon, J., Peterson, A.T., 2009. Monitoring biodiversity loss with primary species-occurrence data: toward national-level indicators for the 2010 target of the convention on biological diversity. AMBIO: A Journal of the Human Environment 38, 29–34.

Sorokina, D., Caruana, R., Riedewald, M., 2007. Additive Groves of Regression Trees, in Machine Learning: ECML 2007, pp. 323–334.

Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D., Kelling, S., 2009. EBird: a citizen-based bird observation network in the biological sciences. Biological Conservation 142, 2282–2292.

Trautmann, N.M., Fee, J., Kahler, P., 2012. Flying into inquiry. Science Teacher 79, 45–50.

Trautmann, N.M., McLinn, C., 2012. Using online data for student investigations in biology and ecology, pp. 80-100 (Chapter 5). In: A.D. Olofsson, O. Lindberg (Eds.), Informed Design of Educational Technologies in Higher Education: Enhanced Learning and Teaching. London: IGI Global.

Tulloch, A.I.T., Possingham, H.P., Joseph, L.N., Szabo, J., Martin, T.G., 2013. Realising the full potential of citizen science monitoring programs. Biological Conservation 165, 128–138.

Western Hemisphere Shorebird Reserve Network, 2003. Sacramento Valley WHSRN Site Profile. <http://www.whsrn.org/site-profile/sacramento-valley>.

Wiggins, A., 2012. Crowdsourcing Scientific Work: A Comparative Study of Technologies, Processes, and Outcomes in Citizen Science, in Information Science and Technology, Syracuse University: Syracuse.

Wiggins, A. 2013. Free as in puppies: compensating for ict constraints in citizen science, in Proceedings of the 2013 conference on Computer supported cooperative work, ACM: San Antonio, Texas, USA. p. 1469–1480.

Wood, C., Sullivan, B., Iliff, M., Fink, D., Kelling, S., 2011. EBird: engaging birders in science and conservation. PLoS Biology 9, e1001220.

Yu, J., Wong, W.-K., Hutchinson, R., 2010. Modeling experts and novices in citizen science data for species distribution modeling. In: Proceedings of the 2010 IEEE International Conference on Data Mining, pp. 1157–1162.