Digital Accessible Knowledge of the birds of India: characterizing gaps in time and space

A. Townsend Peterson, R. Suresh Kumar, Manoj V. Nair and Gautam Talukdar*

This paper evaluates Digital Accessible Knowledge on occurrence of Indian bird species. More than 2 million primary occurrence records from across India were obtained from the Global Biodiversity Information Facility and eBird. These were processed into maps of inventory completeness across the country both prior to 1980 and after 2000, in an attempt to develop evaluations of faunal change resulting from global climate change. We found good coverage of the country by well-inventoried areas after 2000, but almost no coverage prior to 1980. As such, in before-and-after comparisons documenting effects of global change on Indian birds, the 'after' is well documented, but the 'before' is lacking. This significant information gap points to the need for digital capture and open sharing of historical information regarding Indian bird species' occurrences; this information will derive in large part from natural history museum specimens, particularly in India and Great Britain, and potentially from older observational data sources and the literature.

Keywords: Birds, digital accessible knowledge, global change, India, primary biodiversity data.

INDIA is endowed with a rich and highly endemic biota, but the burgeoning human population has massively altered Indian landscapes[1]. The degree to which such widespread human presence compromises the integrity of Indian biotas, depends on the degree to which representative areas remain without disturbance, and on the ubiquity of disturbance across ranges of species. Global climate change effects are particularly relevant as regards the last point. Recent studies in Mexico[2] have illustrated that factors driving biodiversity change can be identified, even across hyperdiverse and environmentally complex tropical landscapes. To this end, it is necessary to marshal large-scale biodiversity information resources and develop comparisons of distributional patterns of species before and after disturbance events[3].

This study represents an assessment of the feasibility of such analyses for the birds of India. The reality of the situation is both optimistic and pessimistic; rich information exists for recent years and many places, thanks to citizen scientists who have generously contributed to biodiversity data sharing initiatives (ebird.org/india). However, in spite of detailed summaries of Indian bird diversity[4], major knowledge gaps remain, both over space and through time, leaving important challenges yet to be met for the country.

A. Townsend Peterson is in the Biodiversity Institute, The University of Kansas, Lawrence, Kansas 66045, USA; R. Suresh Kumar is in the Department of Endangered Species Management, Wildlife Institute of India, Dehradun 248 001, India; Manoj V. Nair is in the UNESCO Category 2 Centre, Wildlife Institute of India, Dehradun 248 001, India and Gautam Talukdar is in the Protected Area Network, WL Management and Conservation Education Department, Wildlife Institute of India, Dehradun 248 001, India.
*For correspondence. (e-mail: town@ku.edu)

This paper, in effect, represents an evaluation of the Digital Accessible Knowledge (DAK)[5] available for Indian bird faunas. We focus explicitly on two time periods – before 1980 and after 2000 – to permit detailed comparisons of avifaunas before and after two decades of global climate change. Our results illustrate important challenges in the field of biodiversity informatics – that is full mobilization of biodiversity information to permit answering of important questions[6–8].

Methods

Occurrence data

We used data from two biodiversity information portals to accumulate data for this set of analyses. First, we obtained data from the Global Biodiversity Information Facility (GBIF; doi:10.15468/dl.luy6wd and doi:10.15468/dl.fztlkh, accessed 15 February 2016). GBIF represents a portal via which data can be obtained from 87 distinct sources of occurrence for Indian birds (Appendix 1). Later in the study, upon learning that GBIF's 2015 list does not include the massive eBird data resource, at least for Indian records, we obtained the November 2015 version of eBird which totalled 2,174,780 records. Given that the GBIF data are not completely up to date, it may be better to download directly from eBird. All records were imported into Microsoft Access for analysis and exploration.

In the initial download of GBIF records, 1,141,588 records corresponded to birds, of which 1,053,547 records included a species-level identification; 1,085,546 had full information on year, month, and day; and 1,076,626 included geographic coordinate information. A total of 988,759 records included all three of these attributes. Gaps in temporal information were found in 54,173 records for which year information was lacking; another ~100 had dates that were well below 1800 and were deemed erroneous. In all, 85.5% of records with year information came after 2000. We explored 901,665 records for which full time, place, and taxon information was available. We also separated records from before 1980 (7.7% of total), totalling 84,146 records initially; however, only 16,690 of these records had complete time-place-taxon data.

Nomenclature

A major challenge in analyses, such as this one, is that of harmonizing species names. Previous analysis showed that uncareful use of names from databases shared online can reduce completeness calculations by at least 3–5% (ref. 5). As a consequence, we extracted all unique names associated with the occurrence data. In all, 1519 names appeared in the data resources. We compared these names with a series of taxonomic authority lists for birds[9–11] and the eBird/Clements checklist[12] (v2015; http://www.birds.cornell.edu/clementschecklist/download/). Matching rates were 1084 species and 730,090 records for Peters, 1145 species and 765,692 records for Morony-Bock-Farrand, 1213 species and 775,034 records for Sibley and Monroe, and 1277 species and 899,872 records from the eBird/Clements list. The high matching is a consequence of the dominance of eBird records in the Indian records. We isolated the taxon names that did not match from all recent records and figured out their identity under 'Clements 2015' taxonomic arrangements using nomenclatural resources in Avibase, and created a correction table that would bring the DAK records into full consistency with the eBird/Clements taxonomy.

This nomenclatural cleaning process involved significant effort. The easiest part was removal of a few records corresponding to higher taxa not represented even in the same hemisphere as India (e.g. *Momotus*, *Seiurus*, *Dendroica*). For more subtle error detection, we downloaded a list of the birds of India in 'Clements 2015' nomenclature (https://goo.gl/71yQ5j), and focused our attention on species taxa recognized as valid by the Clements 2015 list, but not represented in India. In all, we fixed 3829 records in the pre-1980 data set, and 32,039 records in the after-2000 data set. Therefore, the post-2000 dataset consisted of 901,658 records of 1151 species, and the pre-1980 dataset included 16,510 records of 1021 species. Given our focus on terrestrial systems, and the relative paucity of marine records from the region, we further removed marine-bird families and subfamilies (Stercorariidae, Phaethontidae, Hydrobatidae, Procellariidae, Fregatidae and Sulidae), leaving 900,975 records of 1091 species for post-2000, and 15,286 records of 957 species for pre-1980. Even though eBird data quality is examined by a team of reviewers, we note that our taxonomic data cleaning efforts remain challenged by problematic or incorrect identifications among species occurring within India, which are difficult to detect by simple comparison of lists of species recorded versus species known from the country.

Final data preparation

The analyses that are the focus of this study are based on calculations of inventory completeness. They require basic data on place, species identification and time (raw data available at http://hdl.handle.net/1808/24399). The latter was defined as the day (and month and year) on which the record was registered. We created a concatenation of Year_Month_Day, as a time marker for these analyses. Removing the unique identifiers (e.g. GBIF's GBIFID field), we used the unique combinations of time, place (initially latitude and longitude combinations, but later reduced further as data were aggregated spatially), and species identification, which reduced the data set to 700,374 records of 1091 species post-2000, and 12,724 records of 957 species for pre-1980.

Spatial analyses

Spatial exploration of the dataset began with further data cleaning. Simple plotting of the dataset in reference to geographic base maps showed that all of the post-2000 records fell in India, which reflects spatial filtering in the eBird facility, such that a search on India yields only records with coordinates falling in India. Inspection of the pre-1980 data, however, showed a number of points in the North Sea, in Egypt, in North and Central America, and in the Gulf of Guinea. Some of these erroneous points reflected common errors: (1) the points in the Gulf of Guinea were those that had been assigned '0, 0' as geographic coordinates (as an indicator of missing data); (2) the Egyptian point had the same latitude and longitude, suggesting a typographic error; and (3) the North Sea points had latitude and longitude switched. After fixing these problems to every extent possible, we further removed points falling outside of the country's boundaries, to avoid either offshore records of terrestrial taxa or more subtly erroneous geographic coordinates. In the end, these spatial data-cleaning steps reduced the pre-1980 data to 12,487 records and the post-2000 data to 699,323 records.

In the next step, data was aggregated from point-locality geographic information to a grid-based summary that could combine information from nearby localities. We created 'fishnets' (networks of square polygons covering the entire country) for India at resolutions of 0.1°, 0.25°, 0.5°, 0.75°, 1.0°, 1.5° and 2.0°, and reduced each fishnet to just the squares that overlap India. The grid cells at these spatial resolutions ranged in area from 121 to 48,400 sq. km and in number from 28,997 down to 117, across the span of resolutions, from fine to coarse. Point data were aggregated to these grid cells in QGIS; an important point is that, at the finest resolution, for the post-2000 dataset, 25,496 of 28,997 cells (87.9%) held no bird occurrence data, whereas at the coarsest resolution, only 17 of 117 cells (14.5%) held no data.

We explored the density of records in the cells at different resolutions (Figure 1), and – in the end – decided on 0.5° resolution as a best compromise between avoiding aggregation over too-broad areas and avoiding a picture in which the entire region appears devoid of records[13]. To maximize the utility of these calculations, we aligned the fishnet with integer values of latitude and longitude, that creates a direct correspondence between our analysis grid squares and the Survey of India toposheet map series. Finally, upon noting that the southern Indian state of Kerala was particularly densely sampled, we explored finer spatial resolutions in this subregion – we had 125,321 records from this region, and developed completeness evaluations for this state at 0.33° spatial resolution.

Completeness calculations

We based all of our inventory completeness analyses on tables that held the object ID from the 1° fishnet, the time marker described above, and the species identification. Reducing this table to unique combinations yielded a pre-1980 sample size of 12,406 records, and a post-2000 sample size of 553,572 records. We calculated the number of records available (m) for each grid cell, the number of species known (S_{obs}) from each grid cell (i.e. the set of species for which actual records exist from the grid cell), and the numbers of species known from one (Q_1) and two (Q_2) daily records (i.e. species that are known from the grid cell from a record on a single day or on exactly two days) from each grid cell. From these quantities, we calculated the expected number of species using the following equation[14].

$$\hat{S}_{\text{Chao2}} = S_{\text{obs}} + \left(\frac{m-1}{m}\right)\left(\frac{Q_1(Q_1-1)}{2(Q_2+1)}\right).$$

We calculated inventory completeness using the following equation[15]

$$C = \hat{S}_{\text{Chao2}}/S_{\text{obs}},$$

and focused on the grid squares with $C \geq 0.9$. To avoid counting grid squares with few records occasionally appearing to be well-inventoried, we placed a further constraint of $m \geq 200$ and $m \geq 2\,\hat{S}_{\text{Chao2}}$.

Analysis of gaps

Finally, we developed a series of summary visualizations designed to highlight where further sampling is needed

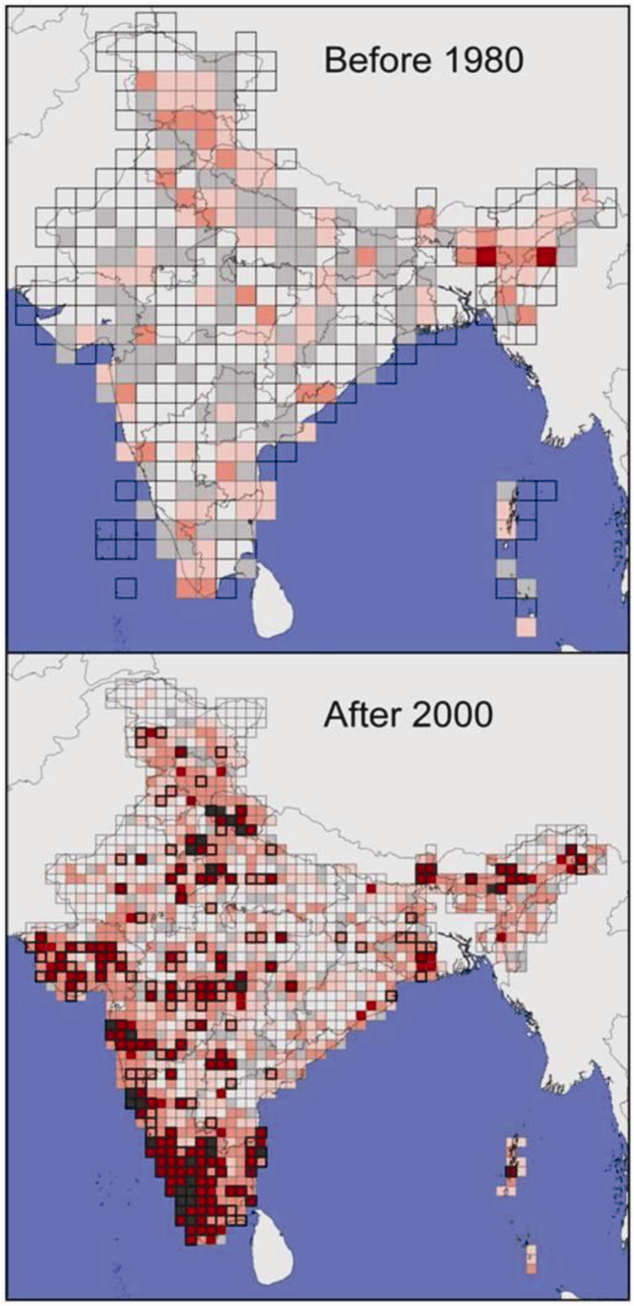

Figure 1. Summary of numbers of records among Digital Accessible Knowledge for the birds of India, and completeness of inventories based on those data. Shading: open = no data, gray = 1–10 records, light pink = 10–100 records, light red = 100–1000 records, red = 1000-10,000 records, black = >10,000 records. Well-known (i.e. $C \geq 0.9$, $m > 200$) squares in the bottom panel are indicated by bold black outlines.

across India. We isolated the grid squares that met two completeness criteria ($C \geq 0.9$ and $m \geq 200$; the latter criterion assures that the high C value is not artefactual), and converted them to raster format. We converted this raster to binary, with well-inventoried grid squares having value 1, and the rest of the region with 0 value. We used the proximity (raster distance) function in QGIS (version 2.12) to summarize geographic distances to well-inventoried sites.

To summarize distance in multivariate climate space to well-inventoried sites, as a metric of climatic difference from well-known sites, we followed methods used in previous analyses[5]. In brief, we used the WorldClim (version 1.4)[16] bioclimatic variables 1–7 and 10–17, at 2.5′ spatial resolution, in a principal components analysis to summarize climatic variation across the country (see visualization of climate variation in Figure 2; we used all 15 components as standard normal variates in analyses). Next, we cast 50,000 random points across the country, and used the point sampling tool in QGIS to add values of environmental variables to each point. We then used the geographic distance raster described above to identify which random points fell in well-inventoried grid squares. Euclidean distance was calculated from all non-well-inventoried grid squares to all well-inventoried grid squares, for extraction of the minimum value. This value was used as the difference in environmental dimensions between the grid square in question and the most-similar well-inventoried grid square.

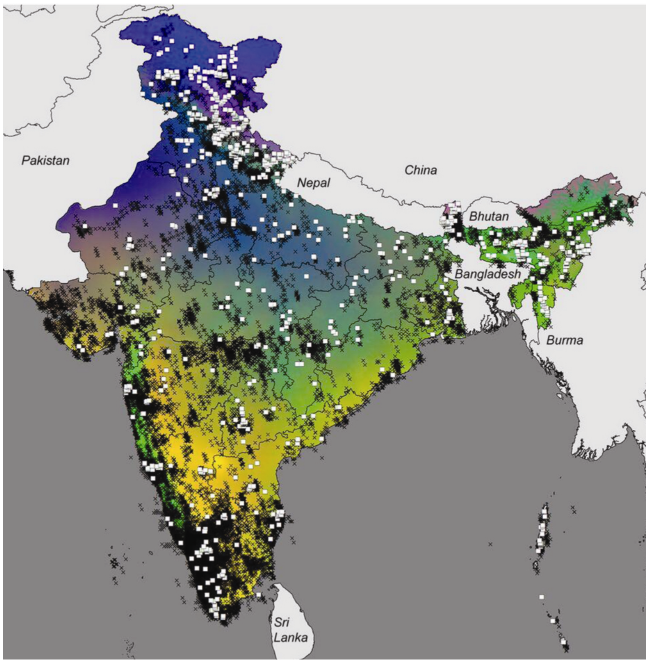

Figure 2. Summary of Digital Accessible Knowledge available for the birds of India. White squares indicate data for before 1980, and X's (black) indicate data for after 2000. The occurrence data are displayed on top of a red-green-blue visualization of the first three principal components of variation in 15 climate dimensions (see text). The difference in colour reflects (roughly) difference in climate. The gray shading represents ocean.

Results

Occurrence data were available from across India, ranging from the western deserts to the easternmost rainforests, and from the tropical south to the Himalaya in the north (Figure 2). We noted concentrations of occurrence data particularly around large metropolitan areas, important and accessible protected areas (PAs), and more generally in Kerala in the deep south. The eastern part of the country has some coverage, although historical (specimen-based) coverage is particularly sparse in the higher-elevation of the Himalayan region of North East India (Figure 2).

Based on post-2000 data, well-inventoried grid squares were distributed across the country, at least in terms of latitude (Figure 1). Well-inventoried grid squares are particularly concentrated in the south, whereas Jammu and Kashmir and the regions disputed with Pakistan and China in the extreme north remain poorly characterized. Curiously, only one grid square is well-inventoried in the eastern third of the country (i.e. east of 82°), such that longitudinal representation is less even. No grid squares qualified as well-inventoried for the pre-1980 period.

Occurrence data were spread evenly and relatively densely across Kerala, so we analysed the region separately, and at a finer spatial resolution (0.33°) (Figure 3). Data were sparse in the interior parts of the north and south of the state. Well-inventoried grid squares numbered 8 out of 51, and were reasonably well scattered across the state, from the south to most of the way north.

Geographic gaps in post-2000 inventory coverage of India were focused in the northwest, north and east. Specifically, gaps were found in desert areas of Rajasthan; across much of Jammu and Kashmir; and in an eastern band across northern Andhra Pradesh, Chattisgarh, Odisha, West Bengal, Jharkhand, eastern Uttar Pradesh and Bihar; and Nagaland, Manipur, Mizoram and Tripura in the north east (Figure 4). Most of the above mentioned areas are primarily less surveyed due to political disturbances. Environmental gaps were focused in the Himalayan north (Jammu and Kashmir), at higher elevations across the Himalayan front, and along the Bangladeshi border.

Discussion

Our results painted a very Dickensian picture of avian inventories across India: the best of inventories and the worst of inventories. That is, thinking about temporal comparisons (i.e. pre-1980 versus post-2000), the post-2000 data were numerous, and were sufficient to characterize avifaunas rigorously for numerous grid squares across the country – that is the good part. On the other hand, however, the earlier data (pre-1980) were painfully sparse, to the point that no sites across the country could be characterized as well-inventoried for that time period.

Gaps in post-2000 coverage

Focusing initially on the post-2000 data, we assessed the gaps in coverage that were detected and characterized particularly PAs including natural habitats. We consider such areas as high priority for inventory, as natural habitats outside of PAs are rare across India. We note that, although some of these sites may genuinely lack surveys, many of them indeed have seen survey efforts, but the data have apparently not been shared openly, so, they cannot be termed as DAK. In a very proximate sense, it may be logical for GBIF to consider quarterly updates of eBird data, at least for countries such as India that can still benefit enormously from additional data resources regarding its biodiversity.

Among the gap areas from Kerala, the most prominent one lies in the eastern portion of the northern district of Kannur, under the hilly Western Ghats portion. The district has one PA, viz. Aralam Wildlife Sanctuary where avifaunal surveys have been conducted and bird watchers do visit, albeit occasionally. However, the results of these surveys and visits are possibly unavailable in the public domain. The information gathered from Kerala illustrates how the quality of geographic summaries mapped from analyses of these data can improve continually, as numbers of records increase.

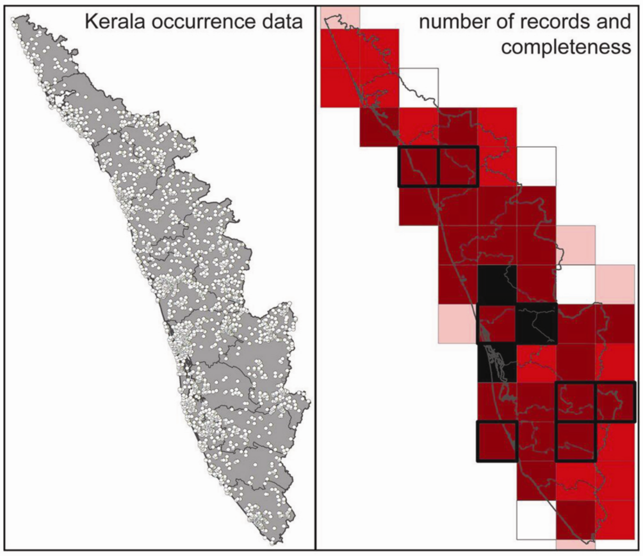

Figure 3. Summary of records among Digital Accessible Knowledge for the birds of the Indian state of Kerala (left panel), and completeness of inventories based on those data (right panel). Shading: open = no data, grey = 1–10 records, light pink = 10–100 records, light red = 100-1000 records, red = 1000–10,000 records, black = >10,000 records, as in Figure 2. Well-known (i.e. $C \geq 0.9$, $m \geq 200$) squares in the right panel are indicated by bold black outlines.

We also note that filling these gaps may prove to be simply a matter of patience. In the 12 months since we completed these analyses, the number of DAK records for Indian birds has apparently increased by 60%. This information is invaluable, as it documents populations of bird species across India, and indeed globally, yet the uptake into global biodiversity information portals is rather slow, and data are not immediately available for science and decision-making. A further point of importance is how credit is appropriately attributed in GBIF.

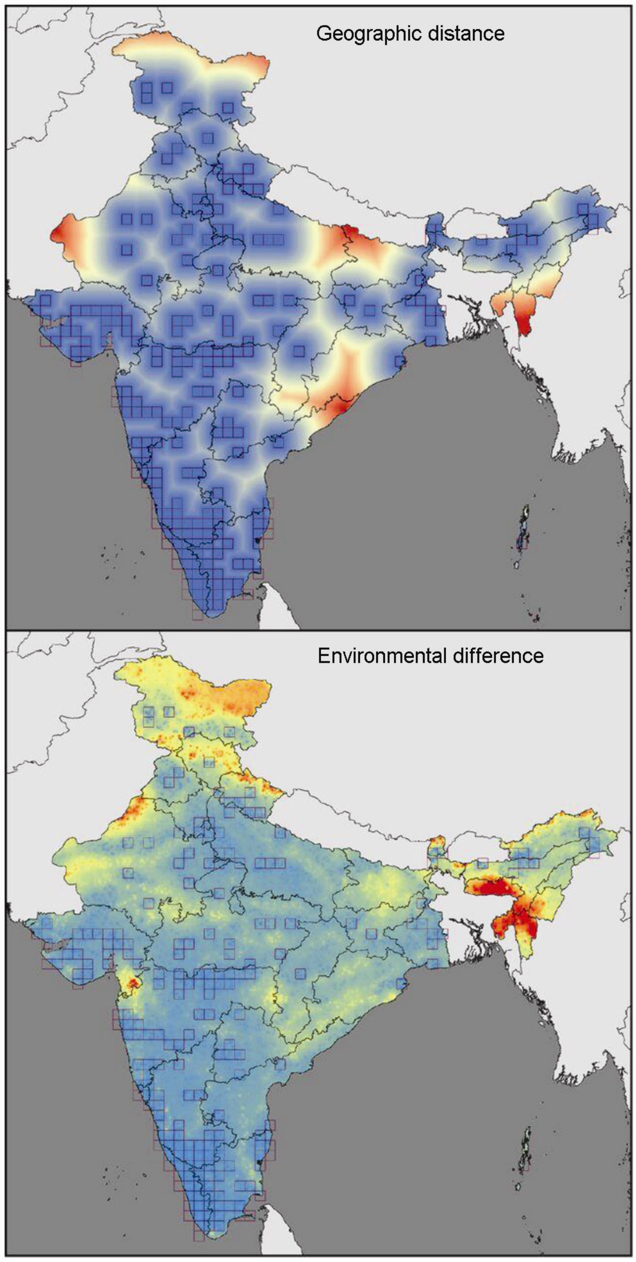

Figure 4. Summary of geographic distance to and environmental difference from a well-inventoried grid square (shown with grey outlines) for the birds of India. Colour ramp extends from blue (close to or similar to well-known squares) through yellow to red (far from or different from well-known squares).

Table 1. Summary of 36 Indian protected areas that emerge as priorities for avifaunal survey, and/or mobilization of avifaunal survey data already in existence

Biogeographic zone	Biotic province	State	Name	Designation
Trans-Himalaya	Ladakh Mountains	Jammu and Kashmir	Hemis	National Park
		Himachal Pradesh	Rakchham Chitkul	Wildlife Sanctuary
Himalaya	West Himalaya	Uttarakhand	Gangotri	National Park
	East Himalaya	Arunachal Pradesh	Tale Valley	Wildlife Sanctuary
Desert	Thar	Rajasthan	Desert	National Park
Semi-Arid	Punjab Plains	Punjab	Abohar	Wildlife Sanctuary
		Punjab	Keshopur Chhamb	Community Reserve
		Chandigarh	Sukhna Lake	Wildlife Sanctuary
		Haryana	Morni Hills (Khol-Hi-Raitan)	Wildlife Sanctuary
	Gujarat-Rajputana	Rajasthan	Mount Abu	Wildlife Sanctuary
Western Ghats	Western Ghats Mountains	Gujarat	Bansda	National Park
Deccan Peninsula	Central Highlands	Madhya Pradesh	Panna (Gangau)	Tiger Reserve
	Chotta-Nagpur	Jharkhand	Dalma	Wildlife Sanctuary
		Odisha	Similipal	Tiger Reserve
	Eastern Highlands	Chhattisgarh	Achanakmar	Tiger Reserve
		Odisha	Chandaka-Dampara	Wildlife Sanctuary
		Odisha	Kotagarh	Wildlife Sanctuary
	Central Plateau	Maharashtra	Andhari	Tiger Reserve
		Maharashtra	Tadoba	Tiger Reserve
	Deccan South	Andhra Pradesh	Sri Lankamalleswara	Wildlife Sanctuary
Gangetic Plain	Lower Gangetic Plain	Bihar	Valmiki	Tiger Reserve
		West Bengal	Buxa	Tiger Reserve
		Uttar Pradesh	Kaimur	Wildlife Sanctuary
		Uttar Pradesh	Ranipur	Wildlife Sanctuary
		Uttar Pradesh	Bakhira	Wildlife Sanctuary
Coasts	East Coast	Andhra Pradesh	Coringa	Wildlife Sanctuary
North-east	Brahmaputra Valley	Assam	Hollongapar-Gibbon	Wildlife Sanctuary
	North-east Hills	Manipur	Keibul-Lamjao	National Park
		Meghalaya	Balphakram	National Park
		Mizoram	Dampa	Tiger Reserve
		Mizoram	Phawngpui (Blue Mountain)	National Park
		Nagaland	Intanki	National Park
		Tripura	Sepahijala	Wildlife Sanctuary
Islands	Andaman	Andaman and Nicobar	Barren Island	Wildlife Sanctuary
		Andaman and Nicobar	Interview Island	Wildlife Sanctuary
		Andaman and Nicobar	Saddle Peak	National Park
	Nicobar	Andaman and Nicobar	Campbell Bay	National Park

The bulk of Indian eBird data comes from Indian observers, largely impelled by the Bird Count India partnership, an online and on-ground effort to mobilize birders toward the goal of generating publicly available occurrence data. Accordingly, as of December 2017, all eBird data from India on GBIF are credited as having been published by India, which is a very welcome step.

Mobilization of data through GBIF is now used by the Convention on Biological Diversity as an official indicator of progress towards *Aichi* Biodiversity Target 19, on sharing of knowledge and biodiversity data. Since these indicators are increasingly disaggregated to national level, excluding such a significant proportion of 'data from' India has the effect of greatly understating India's progress regarding this target. More basically, because eBird is logically and understandably chosen as the most straightforward means of mobilizing observational data on birds, with the understanding that the data will be shared globally, the local and national mobilization efforts involved tend to get obscured.

We identified 37 of 534 Indian PAs that are included in a recent compilation (National Wildlife Database, based at Wildlife Institute of India) but have poor information about its avifauna or there is poor documentation of their avifauna. Most of these sites, at least based on our understanding, are popular areas that bird observers do visit; almost all of these sites are accessible with good road connectivity, and as such their avifauna data should exist. These key sites fall in all 10 Indian biogeographic zones and 18 of the 26 Indian biotic provinces[17,18];

Appendix 1. Summary of sources of digital accessible knowledge regarding Indian birds

Institution	Number of records	Institution	Number of records
Cornell Laboratory of Ornithology/Bird Count India	1,001,248	Museum of Southwestern Biology	44
Wildlife Institute of India	31,936	Western Australian Museum, Perth, Australia	42
University of Michigan Museum of Zoology	30,412	Queensland Museum	36
American Museum of Natural History	17,087	University of Washington Burke Museum	30
Field Museum of Natural History	17,046	Delaware Museum of Natural History	25
Natural History Museum (Tring)	7205	University Museum, Norwegian University of Science and Technology	24
Museum of Comparative Zoology, Harvard University	5392		
United States National Museum of Natural History	4878	California Academy of Sciences	22
Yale Peabody Museum	4410	Illinois State Museum	21
Royal Ontario Museum	3354	Louisiana State University Museum of Natural Sciences	21
Naturgucker	3235	Museum für Naturkunde, Berlin	20
Bombay Natural History Survey	2412	Borror Laboratory of Bioacoustics, Ohio State University	18
Academy of Natural Sciences/Drexel University	2329	Texas Cooperative Wildlife Collections	12
iNaturalist	2229	Musée George Sand et de la Vallée Noire	11
University Museum of Zoology Cambridge	2143	Ohio State University	11
Naturalis	2084	Sam Noble Oklahoma Museum of Natural History	10
Western Foundation of Vertebrate Zoology	1905	Natural History Museum, University of Tartu, Estonia	10
Los Angeles County Museum of Natural History	1316	Moore Laboratory of Zoology	9
Natural History Museum, Oslo, Norway	1194	Anymals.org	8
Royal Ontario Museum: ROM	1081	San Diego Natural History Museum	7
Senckenberg Museum, Frankfurt	999	University of California Los Angeles	7
Zoological Museum Amsterdam	590	Chicago Academy of Sciences	6
Royal Belgian Institute of Natural Sciences	554	Bell Museum of Natural History, University of Minnesota	6
Museum National d'Histoire Naturelle, Paris	529	Wildlife Sightings	6
Cornell University Museum of Vertebrates	456	Zoologische Staatssammlung, Munich	5
Museum and Institute of Zoology, Polish Academy of Sciences	446	Institute of Systematics and Evolution of Animals, Polish Academy of Sciences	4
South Australian Museum, Adelaide	440	James R. Slater Museum, University of Puget Sound	4
Museum of Vertebrate Zoology	321	University of Alaska Museum	4
Santa Barbara Museum of Natural History, Santa Barbara, California	234	Museum of Natural History, University of Colorado	4
		University of Arizona	3
Carnegie Museum of Natural History	227	Angelo State Natural History Collections	2
University of Kansas Natural History Museum	216	Museu de Ciències Naturals de Barcelona	2
National Chemical Laboratory, Pune	215	Estonian Museum of Natural History	2
Michigan State University	190	Überseemuseum, Städtisches Museum, Bremen	2
National Institute of Genetics, ROIS	110	University of Nebraska State Museum	2
Rijksmuseum van Natuurlijke Historie, Leiden	94	Western New Mexico University	2
Lund Museum of Zoology	87	Albany Museum, South Africa	1
Zoologisches Museum der Universitaet Kiel	82	Gothenburg Natural History Museum	1
Yamashina Institute for Ornithology	72	Instituto de Biologia, Universidad Nacional Autonoma de Mexico	1
Provincial Museum of Alberta	71		
Museum Victoria, Melbourne, Australia	69	Indiana State University, Terre Haute, Indiana	1
University of Cape Town Animal Demography Unit	56	German Research Center for Biotechnology	1
Australian National Wildlife Collection	54	North Carolina Museum of Natural Sciences	1
University of British Columbia Beaty Biodiversity Museum	53	Staatliches Museum für Naturkunde, Stuttgart	1
Denver Museum of Nature and Science	47	Tasmanian Museum and Art Gallery	1

they also include 8 tiger reserves. These key sites are summarized in Table 1. Important bird areas, which are increasingly perceived as priority areas for future protection, can also be detailed with respect to completeness of documentation of avifaunas, but we have not as yet been granted a copy of the shapefile that outlines these areas.

Paths forward

The temporal gaps in Indian bird DAK, therefore, are rather massive. As bird observational sampling and reporting are rather new phenomena in India[19], mobilizing large-scale historical DAK resources for Indian birds will depend largely on data from scientific specimens. Some progress can be and has been made via the scientific literature[20]. As these resources grow, an interesting and relevant question will be of how sampling by observers and sampling by specimen collectors differs, and how these different processes translate into different patterns of knowledge and completeness.

The GBIF-mediated data (across all time periods) included 100,238 specimen records, of which only 38,829 had geographic references, and 19,949 lacked adequate temporal information (not even information on year of collection). These data represent the bulk of Indian bird specimen holdings in collections of the University of Michigan (28,356 specimens), American Museum of Natural History (17,008 specimens), and the Field Museum of Natural History (16,083 specimens); from the Natural History Museum (Tring), only 6650 records are available, although the institution's holdings may be more than 20-fold of that amount (P. Rasmussen, pers. comm.). Indian institutions such as the Bombay Natural History Survey and the Zoological Survey of India also hold significant historical collections of Indian bird specimens, which remain to be incorporated into these analyses.

Clearly, more specimen data exist to document Indian bird diversity. Major collections worldwide include the Natural History Museum (Tring), Bombay Natural History Society (Mumbai), and the Zoological Survey of India (Kolkata), which together would likely at least triple the numbers of historical specimens and associated data available from the country. Collecting and sharing data represent an important step for advances in Indian ornithology. Once the great bulk of existing bird occurrence data transitions into being DAK, accessible to the broader community, and integrated with other such data – major progress can be made in terms of (1) understanding Indian bird geography, (2) optimizing conservation strategies, and (3) identifying gaps for on-ground survey and inventory work.

1. Singh, J. S. and Kushwaha, S. P. S., Forest biodiversity and its conservation in India. *Int. For. Rev.*, 2008, **10**, 292–304.
2. Peterson, A. T., Navarro-Sigüenza, A. G., Martínez-Meyer, E., Cuervo-Robayo, A. P., Berlanga, H. and Soberón, J., Twentieth century turnover of Mexican endemic avifaunas: Landscape change versus climate drivers. *Sci. Adv.*, 2015, **1**, e1400071.
3. Peterson, A. T., Navarro-Sigüenza, A. G. and Martínez-Meyer, E., Digital accessible knowledge and well-inventoried sites for birds in Mexico: baseline sites for measuring faunistic change. *Peer J.*, 2016, **4**, e2362.
4. Ali, S. and Ripley, S. D., *Handbook of the Birds of India and Pakistan. Together with those of Bangladesh, Nepal, Sikkim, Bhutan and Sri Lanka. Vol. 3: Stone Curlews to Owls*, Oxford University Press, Bombay, 1969.
5. Sousa-Baena, M. S., Garcia, L. C. and Peterson, A. T., Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. *Divers. Distrib.*, 2013, **20**, 369–381.
6. Peterson, A. T., Knapp, S., Guralnick, R., Soberón, J. and Holder, M. T., The big questions for biodiversity informatics. *Syst. Biodiv.*, 2010, **8**, 159–168.
7. Hardisty, A. and Roberts, D., A decadal view of biodiversity informatics: challenges and priorities. *BMC Ecol.*, 2013, **13**, 16.
8. Peterson, A. T., Soberón, J. and Krishtalka, L., A global perspective on decadal challenges and priorities in biodiversity informatics. *BMC Ecol.*, 2015, **15**, 1–9.
9. Peters, J. L., *Check-list of Birds of the World*, Harvard University Press, Cambridge, 1931–1987, vols 1–16.
10. Morony, J. J., Bock, W. J. and Farrand, J., *Reference List of the Birds of the World*, American Museum of Natural History, New York, 1975.
11. Sibley, C. G. and Monroe, B. L. J., *Distribution and Taxonomy of Birds of the World*, Yale University Press, New Haven, Connecticut, 1990.
12. Clements, J. F., Schulenberg, T. S., Iliff, M. J., Roberson, D., Fredericks, T. A., Sullivan, B. L. and Wood, C. L., The eBird/Clements checklist of birds of the world, 2015.
13. Lira-Noriega, A., Soberón, J., Navarro-Sigüenza, A. G., Nakazawa, Y. and Peterson, A. T., Scale-dependency of diversity components estimated from primary biodiversity data and distribution maps. *Divers. Distrib.*, 2007, **13**, 185–195.
14. Colwell, R. K., 1994–present. EstimateS. University of Connecticut, Storrs; http://viceroy.eeb.uconn.edu/estimates/
15. Peterson, A. T. and Slade, N. A., Extrapolating inventory results into biodiversity estimates and the importance of stopping rules. *Divers. Distrib.*, 1998, **4**, 95–105.
16. Hijmans, R., Cameron, S., Parra, J., Jones, P. and Jarvis, A., Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.*, 2005, **25**, 1965–1978.
17. Rodgers, W. A. and Panwar, H. S., *Planning a Wildlife Protected Area Network in India*, Wildlife Institute of India, Dehradun, 1988, vols 1 and 2.
18. Rodgers, W. A., Panwar, H. S. and Mathur, V. B., Wildlife Protected Area Network in India: A Review (Executive Summary). Wildlide Institute of India, Dehradun, 2002.
19. Urfi, A. J., Sen, M., Kalam, A. and Meganathan, T., Counting birds in India: methodologies and trends. *Curr. Sci.*, 2005, **89**, 1997–2003.
20. Narwade, S. *et al.*, Literature based species occurrence data of birds of northeast India. *ZooKeys*, 2011, **150**, 407–417.

ACKNOWLEDGEMENTS. We thank the leaders of eBird India and eBird (international) for their help. We are grateful to the natural history museums for providing their Indian bird data. We also thank the U.S. Fulbright Specialists Program, for funding ATP's extended stay at the Wildlife Institute of India, during which time this work was developed. Department of Biotechnology IBIN project is acknowledged for providing support.

Received 16 February 2017; revised accepted 22 February 2018

doi: 10.18520/cs/v115/i1/35-42