www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN

DATA DESCRIPTOR

GalliForm, a database of Galliformes occurrence records from the Indo-Malay and Palaearctic, 1800–2008

Elizabeth H. Boakes[1]✉, Richard A. Fuller[2], Georgina M. Mace[1], Changqing Ding[3], Tzo Tze Ang[4], Alistair G. Auffret[4,5], Natalie E. Clark[4,6], Jonathon Dunn[7], Jennifer Gilbert[4], Viktor Golovnyuk[8], Garima Gupta[7], Ulrike Irlich[4,9], Emily Joachim[4], Kim O' Connor[4], Eugene Potapov[10], Roald Potapov[11,14], Judith Schleicher[4,12], Sarah Stebbing[4], Terry Townshend[13] & Philip J. K. McGowan[7]

Historical as well as current species distribution data are needed to track changes in biodiversity. Species distribution data are found in a variety of sources, each of which has its own distinct bias toward certain taxa, time periods or places. We present GalliForm, a database that comprises 186687 galliform occurrence records linked to 118907 localities in Europe and Asia. Records were derived from museums, peer-reviewed and grey literature, unpublished field notes, diaries and correspondence, banding records, atlas records and online birding trip reports. We describe data collection processes, georeferencing methods and quality-control procedures. This database has underpinned several peer-reviewed studies, investigating spatial and temporal bias in biodiversity data, species' geographic range changes and local extirpation patterns. In our rapidly changing world, an understanding of long-term change in species' distributions is key to predicting future impacts of threatening processes such as land use change, over-exploitation of species and climate change. This database, its historical aspect in particular, provides a valuable source of information for further studies in macroecology and biodiversity conservation.

Background & Summary

Gathering primary biodiversity data is necessary to improve our knowledge of the ecology and conservation status of species. International commitments such as the Convention on Biological Diversity[1] call for a halt to biodiversity loss and therefore require data to measure biodiversity change. Recent trends in changes in population sizes or geographical ranges can be used to track progress toward biodiversity targets but longer-term trends are needed if we are to put the status of present-day biota into a proper historical context[2,3]. Similarly, if we are to understand the impacts of climate and land use change on species distributions, historical data are required. Ideally, this biodiversity information must be comprehensive, covering common species as well as threatened, and areas of lower biodiversity as well as hotspots.

[1]Centre for Biodiversity and Environment Research, University College London, Gower Street, London, WC1E 6BT, UK. [2]School of Biological Sciences, University of Queensland, Brisbane, QLD 4072, Australia. [3]College of Biological Sciences and Biotechnology, Beijing Forestry University, Beijing, 100083, China. [4]The Institute of Zoology, Zoological Society of London, Regents Park, London, NW1 4RY, UK. [5]Department of Ecology, Swedish University of Agricultural Sciences, Box 7044, 75007, Uppsala, Sweden. [6]National Environment Research Council, Polaris House, North Star Avenue, Swindon, SN2 1EU, UK. [7]School of Biology, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK. [8]FSBI "Taimyr Reserves", Talnakhskata str 22, Norilsk, 663305, Russia. [9]Fisheries and Oceans Canada, 200 Kent St, Ottawa, Ontario, Canada. [10]Department of Biology, Bryn Athyn College, 2945 College Drive, Bryn Athyn, PA, 19009, USA. [11]Zoological Institute, Russian Academy of Sciences, St Petersburg, 199034, Russia. [12]Department of Geography, University of Cambridge, Cambridge, CB2 1QB, UK. [13]Birding Beijing, Great Walsingham, Norfolk, NR22 6DR, UK. [14]Deceased: Roald Potapov. ✉e-mail: e.boakes@ucl.ac.uk

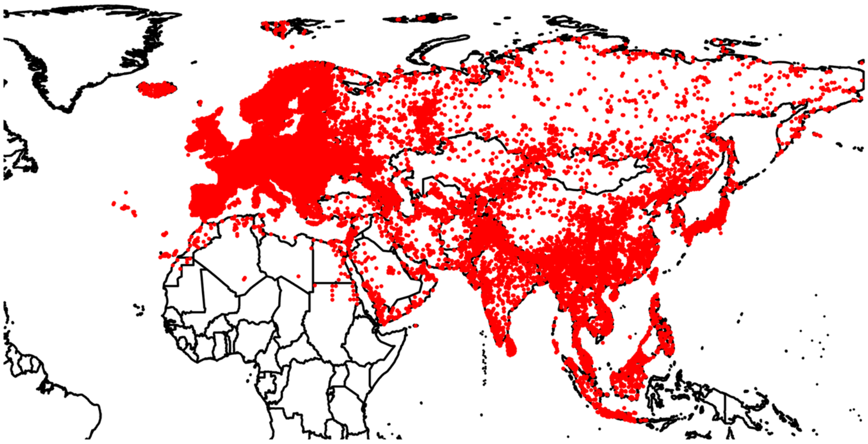

Fig. 1 The spatial distribution of those records in GalliForm that contain sufficient information to be georeferenced to an accuracy of 30 minutes. The records of *Lagopus lagopus* and *Lagopus muta* from North America are omitted.

Our knowledge of species' distributions is extremely coarse compared to most other environmental variables[4]. Analyses of species' geographical ranges often rely on predictions of where a species might occur. Predictions might be gleaned from expert opinion (e.g. https://birdsoftheworld.org/bow/home) (and in some instances may be influenced by historical data), the extent of suitable habitat[5], gridded survey data[6] or point occurrences[7]. Prominent conservation datasets such as the Living Planet Index[8] and IUCN's species distribution maps (https://www.iucnredlist.org/resources/spatial-data-download) are regularly used to assess rates of biodiversity loss but these data sources do not extend back beyond around 1970. Longer-term trends can reveal major shifts in abundance and composition of biological communities, information that should be considered when setting conservation targets[9].

While aggregated population trends or extent of occurrence maps are useful conservation tools, primary data allow us to investigate biodiversity loss in far greater detail. For example, if species' ranges are punctuated with local extinction events we might overlook or underestimate species' declines because we lack the precision to measure them[10]. Additionally, data summaries may be at coarser resolutions than the original data or missing attributes attached to the original record. Freely available primary data allow new questions to be investigated, for which data summaries might not be suitable.

The avian order Galliformes has relatively high quality historical distribution data. This is in part due to their economic and cultural value and their attraction for collectors and ornithologists[11]. Almost all species are non-migratory, making delimitation of their current and historical ranges more tractable. In recent times they have received much conservation attention through being one of the most threatened avian orders – over 25% of species are threatened (www.iucnredlist.org) and many local extinctions have been reported[12] (http://datazone.birdlife.org/home). Galliformes are subject to a variety of threats including habitat loss, hunting, and agricultural intensification and disturbance (http://datazone.birdlife.org/home). The order exhibits a wide range of ecological characteristics and life history traits, and occurs in a diversity of habitats, meaning that the Galliformes lend themselves well to macroecological studies[13].

Here we present GalliForm[14], a database of 186687 occurrence records covering the 130 species of the avian order Galliformes that occur in the Palaearctic and Indo-Malay biogeographic realms (see Fig. 1 for spatial distribution of records). Records cover the period 1648 to 2008 although 95% of records date from 1877 onwards. Records increase markedly though time (Fig. 2). Records were collected from museums, peer-reviewed and grey literature, bird atlases, banding records and birding trip report websites (see[15] for spatial biases within sources). Where possible, data were informally refereed by local experts who, if necessary, supplemented the data with their personal records. Each data source was found to have a distinct set of spatial, temporal and taxonomic biases[15]. Combining biodiversity data from a variety of primary sources helps to minimise data bias.

The GalliForm dataset[14] is an extremely valuable resource for ecological and conservation studies. Occurrence data underpin species distribution modelling but geographic ranges are changing rapidly due to the diverse impacts caused by human activities. Historical occurrence data, coupled with climate and land-use data, may improve our understanding of populations' responses to climate change, land-use change and hunting. The species occurrence data described here have been used to assess the completeness of geographic range size estimates[16], to investigate patterns of range collapse with respect to distance to range edge[17] and to assess species extirpations outside Protected Areas[12]. Nine publications[10,12,15–21] have so far arisen from this database but many avenues remain to be explored.

Methods

These methods are an expanded version of those in our related work, Boakes *et al.*[15].

The database was compiled over the period 2005–2008. Data collection equates to around 1500 person-days and data were gathered by a team of 21 people. Between them, team members were fluent in English, French, German, Mandarin, Russian, Spanish and Swedish. These languages were extremely helpful in transcribing

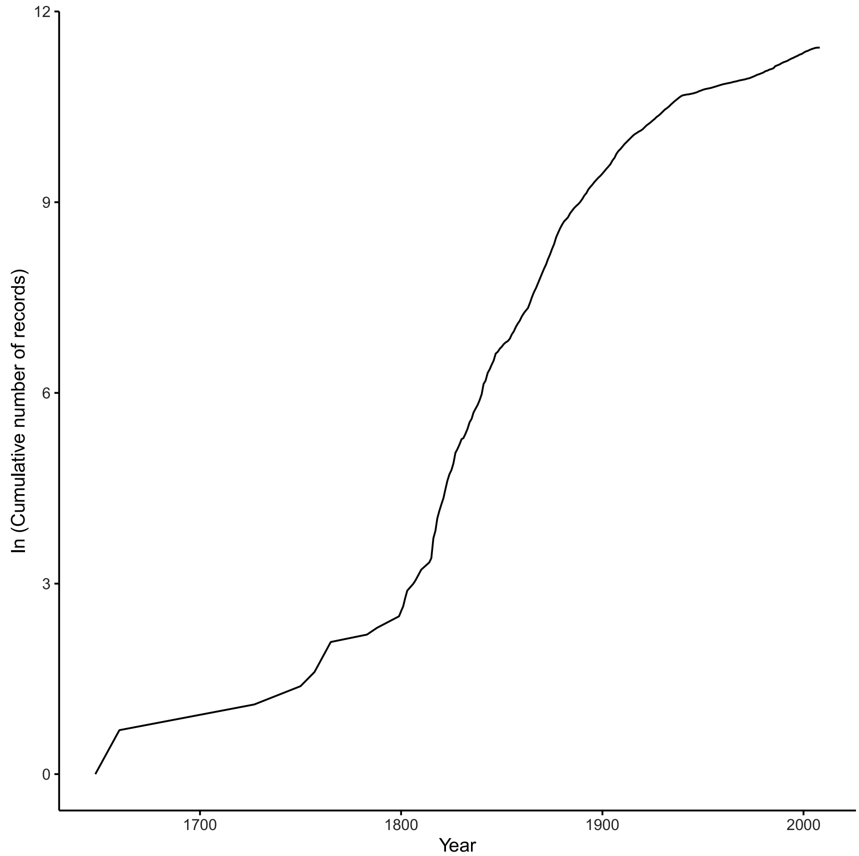

Fig. 2 The cumulative number of occurrence records through time. The number of occurrence records has been converted to a natural logarithmic scale.

museum specimen labels and in translating publications. However, the majority of publications were in English and we acknowledge that the database will be biased toward records published in English-language publications.

Our study focuses on the 130 galliform species that occur within the Palaearctic and Indo-Malay biogeographic realms[22] (see Online-only Table 1). We have additionally included records of the Imperial Pheasant (*Lophura imperialis*) although it is now recognised that this is a hybrid and not a species. The geographic range of two of the species in the database, the Red Grouse (*Lagopus lagopus)* and the Rock Ptarmigan (*Lagopus muta*), extends to North America. North American data was often included in the information which museums sent us and in these instances we entered those records into the database since we thought they might be of use to researchers studying these species. However, it should be noted that we did not search exhaustively for records of these species in North America, we have merely included those that we came across.

We attempted to gather all species distribution data that could be accessed from five different sources; museum collections, literature records, banding (ringing) data, ornithological atlases and birdwatchers' trip report websites. For each data source, exhaustive and systematic search strategies were adopted.

Museum collections. Using web-based searches and Roselaar[23], 377 natural history collections were identified. We found contact details for 338 of these collections and requested by email or letter a list of the Galliformes in their holdings along with collection localities and dates. Non-respondents were recontacted. 135 museums were able to share data with us (see Online-only Table 2). Museum records were obtained through publicly available online databases e.g. ORNIS, electronic or paper catalogues sent to us by the museums or by visiting the museums and transcribing data directly from specimens or card catalogues. Almost half of the museums we contacted did not respond despite at least one follow-up enquiry, and there was substantial variation in the amount and format of data contributed by those that did reply. Altogether, over 50% of the records came from just six museums (Natural History Museum, London; Zoological Institute of the Russian Academy of Sciences, St Petersburg; Zoological Museum of Lomonosov Moscow State University; Field Museum of Natural History, Chicago; American Museum of Natural History, New York; National Museum of Natural History, Leiden), a single museum (the Natural History Museum, London) contributing nearly 20% of the museum records that could be georeferenced and dated[15]. Following databasing and/or georeferencing, records were returned to larger collections and to those who had requested the data.

Literature. Data from the literature were added to those previously collected by McGowan[24]. Entire series of key English-language international and regional ornithological journals such as *Ibis, Bird Conservation*

Ringing group
EURING
Zagreb Ringing Scheme
Hungarian Bird Ringing Centre
Finnish Museum of Natural History, Ringing Centre
Beringungszentrale Hiddensee
Coturnix ringing records, Italy
National Parks Board, Singapore (Ringing Centre)

Table 1. The ringing groups that shared data with GalliForm.

International, Journal of the Bombay Natural History Society, and *Kukila* were scanned for relevant information, availability allowing. We began at the library of the Zoological Society of London and followed up missing journal issues at the BirdLife International library, Cambridge UK; the British Library, London, UK; the Edward Grey Institute, University of Oxford, UK. Relevant Chinese literature was also scanned. Additionally, data were obtained from regional reports, personal diaries, letters, newsletters etc stored in the archives of BirdLife International, Cambridge, UK; the World Pheasant Association, Newcastle, UK; the Edward Grey Institute, University of Oxford, UK. Several of the species/regional experts we consulted also contributed their personal records which were recorded in the database as 'personal communications'. As far as it were possible, records were classed as primary or secondary data within the 'dynamicProperties' field of GalliForm[14]. It is important to note that some primary records or museum specimens will be duplicated within the database in the secondary data.

Banding records. Eighty-three ornithological banding groups were identified using web-based searches and were contacted via email. Thirty of these groups replied and only seven were able to provide us with data (see Table 1). The majority of galliform species tend not to be banded due to their large body sizes and spurs. Additionally, many of the banding groups kept their records on paper and were not able to send them to us. Nevertheless, we were able to access and georeference 15,152 banding records.

Ornithological atlases. We digitised location data from 20 ornithological atlases (see Table 2). Data from several other atlases were not used since the range of dates for the records was wider than 20 years.

Trip report website data. We used the two trip report websites that were popular with birders during the data recording period (2005–2008), www.travellingbirder.com and www.birdtours.co.uk. At that time, eBird (probably the most relevant current online source today) did not cover the majority of the countries within our study region, and our intention with the deposition of this dataset is to focus on pre-eBird data that are more difficult and time consuming to access. We extracted data from all trip reports of birdwatching visits to European, Asian and North African countries. Care was taken to enter reports that featured on both websites once only.

Criteria for data inclusion. To be included in the database, records had to meet the following criteria:

1. The record identified the species of the bird concerned.
2. The record contained either a verbal description of the locality at which the bird concerned was observed or the co-ordinates at which the bird was observed.

Records of captive birds were excluded. Records relating to non-native occurrences were included but were flagged in the 'establishmentMeans' field as "introduced".

Data entry. GalliForm[14] was originally compiled in the programme Microsoft Access 2003. To maximise uniformity in data entry, all data recorders were given thorough and consistent training and each was provided with a set of database guidelines. An Access Database form was created to standardise data entry and to enable multiple members of the team to collect data simultaneously.

Each entry in GalliForm[14] corresponds to a single record of a single species recorded in a specific location. The data fields of GalliForm[14] are described in Online-only Table 3. The taxonomy used has been updated to be consistent with the BirdLife International 2019 taxonomy (datazone.birdlife.org). All information was entered exactly as it was described in the data source, with as much information extracted as possible. Multiple records from different sources which recorded the same information were still included in the interest of completeness. The only exception to this is the trip report data in which we did not enter identical records which occurred on both the Travelling Birder and Bird Tours websites.

The source of the data, i.e. literature, museum, atlas, ringing or website trip report is recorded in the 'dynamicProperties' field under the code "dataSource". For literature data, (where known) the nature of the record, i.e. primary or secondary, is recorded under the code "datatype".

Taxonomy has of course changed considerably over time. To allow for this we recorded the taxonomy as it was described in the data source in the 'originalNameUsage' field. The current taxonomy was then selected from a look-up table. If at the time of data entry, the data compiler was unsure which species the synonym referred to, the species was tagged as "unknown" and the species was designated at a later date following further research on the synonym.

Atlas	Year	Editors
The EBCC atlas of European breeding birds: their distribution and abundance[6]	1997	Hagemeijer, E.J.M. & Blair, M.J.
The atlas of breeding birds in Britain and Ireland[30]	1976	Sharrock, J.T.R.
The new atlas of breeding birds in Britain and Ireland[31]	1993	Gibbons, D.W.
Atlas of breeding birds of the West Midlands[32]	1970	Lord, J., Munns, D.J.
Atlas of the breeding birds of Andorra[33]	2002	Alamany, O., Auclair, R., Bertrand, A.
Atlas des oiseaux nicheurs de Belgique[34]	1988	Devilliers, P., Roggeman, W., Tricot, J., Del Marmol, P., Kerwijn, C., Jacob, J-P., Anselin, A.
Atlas of breeding birds in Luxembourg[35]	1987	Melchior, E.
Atlas van de Nederlandse Broedvogels 1973–1977[36]	1979	Teixeira, R.M.
Atlas van de Nederlandse Broedvogels 1978–1983[37]	1987	
Atlas das aves que nidificam em Portugal Continental[38]	1989	Rufino, R.
Atlante degli uccelli nidificanti e svernanti in Toscana[39]	1997	Florenzano, G.T., Arcamone, E., Baccetti, N., Meschini, E., Sposimo, P.
Atlas Hnizdniho Rozsireni Ptaku V CSSR[40]	1987	Stastny, K., Randik, A., Hudec, K.
Birds of Moscow city and the Moscow region[41]	2006	Kalyakin, M.V., Voltzit, O.V.
Eesti Linnuatlas[42]	1993	Renno, O.
Latvian breeding bird atlas[43]	1989	Priednieks, J., Strazds, M, Strazds, A. and Petrins, A.
Zimski ornitoloski atlas Slovenije[44]	1993	Sovinc, A.
Breeding bird atlas of Oman[45]	1998	Eriksen, J.
An interim atlas of the breeding birds of Arabia[46]	1995	Jennings, M.C.
Distribution atlas of Sudan's birds with notes on habitat and status[47]	1987	Nikolaus, G.
Atlas of wintering birds of Japan[48]	2004	

Table 2. The atlases that were digitised to be included in GalliForm.

Identical localities can also be described in multiple ways. We recorded the locality as it was given in the data source in the 'verbatimLocality' field. If the 'verbatimLocality' clearly tallied with a locality already within the database, the record was linked to that locality in order to increase georeferencing efficiency.

It was rare for a source to record absence of evidence, i.e. a survey for a species at a particular locality which failed to find that species. However, in the few cases where we did come across such records, the locality and date of the survey were recorded and "absent" was recorded in the 'occurrenceStatus' field.

Each record refers to an independent observation. For museum and ringing records, this means a single individual. For literature, atlas or trip report records this may refer to a group of birds observed in one particular locality, on one particular day. If given, the number of total individuals is recorded in the 'individualCount' field. The number of males and females is recorded in the 'sex' field and the number of juveniles and adults in the 'lifeStage' field. If the 'lifeStage' field is blank, it is reasonable to assume the individual(s) is an adult.

Occasionally, additional information about the observation might be included in the data source, for example the habitat the bird was observed in or whether the bird was common or rare in that locality. These data are recorded in the 'habitat' and 'organismQuantity' fields, respectively. Any additional information which did not fit within the structure of the database was recorded in the 'occurrenceRemarks' field, along with any notes found on museum labels.

For the purposes of data deposition, the database was converted to a tab-delimited CSV file with all fields following Darwin Core format. A full summary of these fields is given in Online-only Table 3.

Georeferencing. Locality descriptions were converted to geographic co-ordinates using a wide range of atlases and gazetteers, co-ordinates generally only being assigned if accurate to one degree (although in the majority of cases the locations were accurate to within 30 minutes, Table 3). We would initially search for a locality

Record Class	No. records	No. georeferenced to within 2 minutes	No. georeferenced to within 10 minutes	No. georeferenced to within 30 minutes	No. dated to within one year	No. dated to within 10 years	No. georeferenced to within 30 minutes and dated to within one year
Event	186687	57173 (31%)	58773 (32%)	152930 (82%)	91973 (49%)	165312 (89%)	65913 (35%)
Locality	118907	26282 (22%)	26755 (23%)	109651 (92%)	N/A	N/A	N/A

Table 3. Georeference and date completeness of the records.

within the gazetteers available to us at the time. If the locality was not listed within those gazetteers we would search for the locality using atlases. Since this fieldwork was conducted, MaNIS standards have become widely used for studies of this kind, but these weren't fully developed at the time of data collection[25]. Named places, e.g. towns or counties, were georeferenced using their geographic centre and georeferencing uncertainty measured from the centre to the edge of the named place. Often localities were given simply as the name of a river, mountain or Protected Area. In these instances we used the midpoint of the river between source and mouth (uncertainty measured as distance from midpoint to source/mouth), the summit of the mountain (uncertainty measured as distance from summit to approximate mountain foot) and the rough centre of the Protected Area (uncertainty measured as distance from centre to Protected Area edge). If a particular locality description matched two or more places their midpoint was taken (uncertainty measured as distance from midpoint to place). Offsets from localities (e.g. "50 km N of Kuala Lumpur"; "8 miles along the road from Sheffield to Chesterfield") were measured using a digital atlas (uncertainty was approximated at the georeferencer's discretion in these instances, usually between 3 and 10 arc-minutes, depending on the vagueness of the offset.) For georeferencing done 'in house', the gazeteer/atlas used was recorded.

When possible, localities we could not georeference ourselves were sent to regional experts.

92% of our localities are georeferenced to an accuracy of 30 minutes, corresponding to 82% of occurrence records (see Table 3).

We had less success at georeferencing museum records than literature records[15], due in part to difficulties in reading hand-writing on specimen labels. Older records were also harder to georeference, presumably due to changes in place names over time, and to some early ornithologists failing to document the collection locality. As might be expected, localities from countries that do not use the Roman alphabet were also harder to georeference.

Some records were excluded from the database based on their locality: records which we thought were trading localities, notably Malacca in Malaysia and Leadenhall Market in the UK; records from captive specimens, e.g. zoological gardens.

Dating. 49% of records are dated to within an accuracy of one year. Where possible, we assigned date ranges to undated records. For example, if the name of the collector was given on a museum specimen and we knew when that collector was active in that region, we assigned a date range covering that period. There remain undated records which could perhaps be dated in this way. Undated literature records were designated as occurring before their publication date. We were able to date 89% of records to within 10 years.

Data Records

A relational database structure was created in Microsoft Access to organise and store the species occurrence records with their spatial dependencies and data sources and to keep track of synonyms. For the purposes of publication, this database was converted to a tab-delimited CSV file that followed the Darwin Core format.

We provide a dataset for Galliformes occurrences within the Palaearctic and Indo-Malay realms at species level. These data, obtained and curated as explained above, are available from the Global Biodiversity Information Facility (https://doi.org/10.15468/9825yw). Online-only Table 3 lists and describes the fields of GalliForm[14].

The following figures and tables summarise the dataset. Figure 1 shows the spatial distribution of records; Fig. 2 shows the accumulation of records through time; Fig. 3 shows the spatial distribution of the number of records, species richness and the most recent year of record; Fig. 4 shows the completeness of selected data fields. Table 1 lists the ringing groups which were able to share data with us; Table 2 lists the atlases that we digitised; Table 3 details the completeness of records which are georeferenced and/or dated to within 1 year. Online-only Table 1 details the number of records per species and the time span these records cover; Online-only Table 2 lists the museums which were able to share data with us; Online-only Table 3 describes the Field Names of GalliForm[14].

Technical Validation

Georeferenced data were subject to the following checks:

1. That each data point was in the country that its locality described.
2. That each data point was within reasonable distance of the species' known historical range.
3. That each data point that identifiably came from a protected area listed in the World Database of Protected Areas (https://www.protectedplanet.net/) was indeed within that protected area.

Finally, data were sent to experts on regions/species for informal 'refereeing' to highlight dubious or missing data. We were able to referee approximately one third of the records in this way.

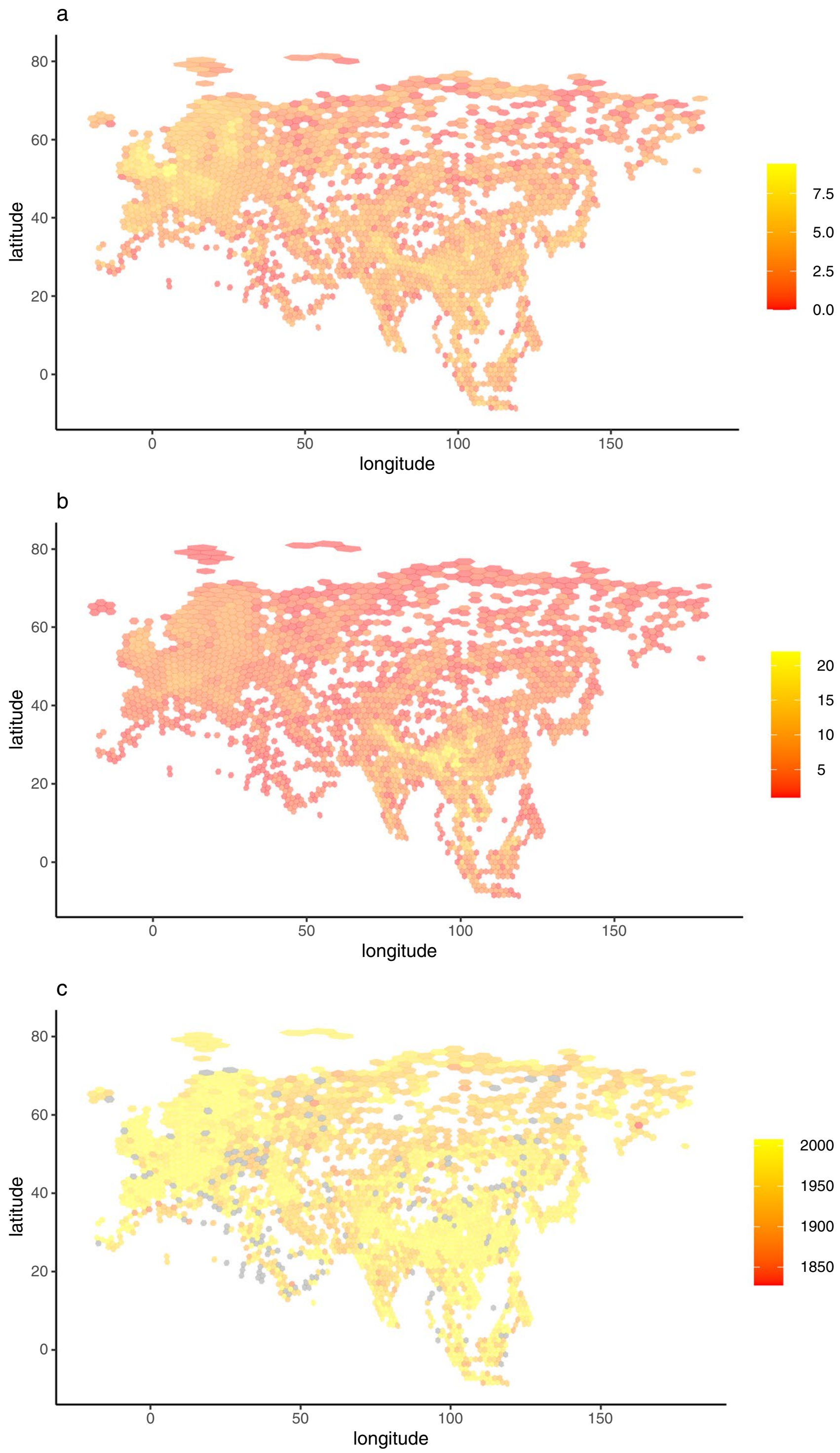

Fig. 3 The spatial distribution of the records, coloured coded by (**a**) the natural logarithm of the number of records within each cell, (**b**) the number of species within each cell and (**c**) the most recent year of record within each cell (cells which do not contain any dated records are shaded light grey). Cells are equal area and represent approximately 23,322 km^2. Cells were drawn using the dgGridR package[28] in R[29].

Usage Notes

The dataset described here can be used to investigate the spatial and temporal patterns of Galliformes distributions at multiple scales and resolutions. The dataset was first used to examine bias in different sources of biodiversity data[15]. It has also been used to investigate predictors of range change[18], to examine the effects of missing data on estimates of biodiversity metrics[10], to assess the completeness of geographic range estimates[16], to investigate the position of local extinctions with respect to species' range edges[17], to explore the optimisation of Protected

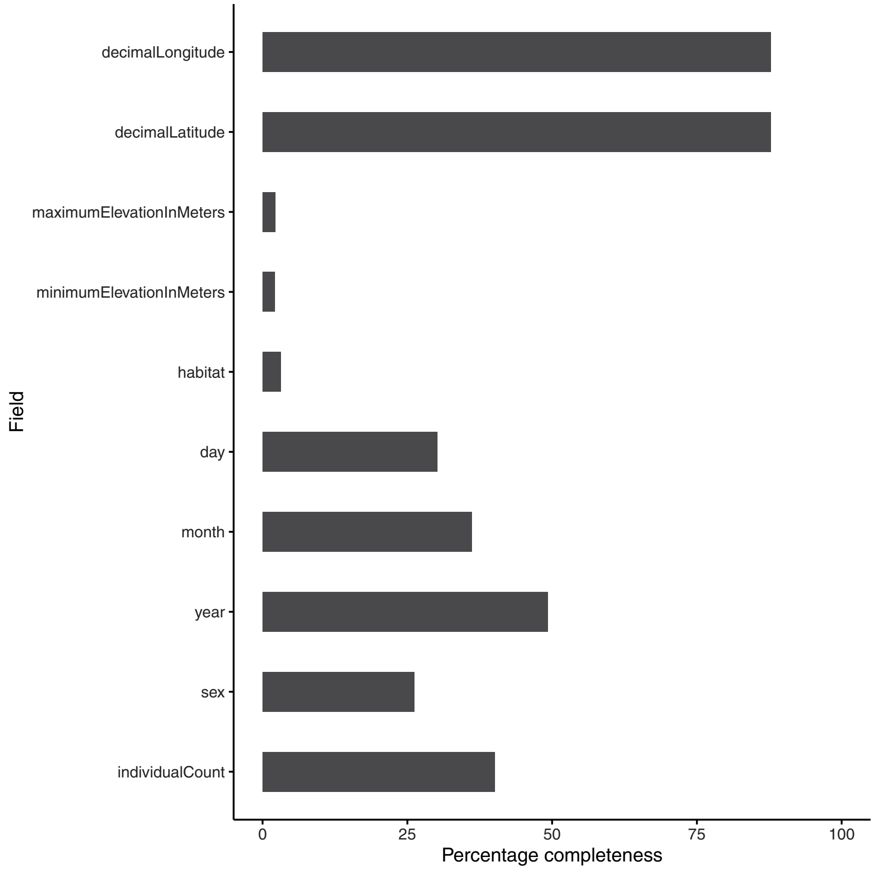

Fig. 4 Percentage of data completeness of selected fields of GalliForm. Field descriptions are given in Online-only Table 3.

Area networks[20], to examine the local extirpation of species outside Protected Areas[12] and to model the potential distributions of highly threatened species[19,21]. There remains much scope for this database to inform further biodiversity or conservation related studies, for example, investigations of geographic range change or predictors of extinction risk.

The data presented here do need to be interpreted carefully with respect to data bias and to missing data. Biodiversity data may be biased in a variety of ways, for example geographically, towards particular ecosystems or towards more charismatic species e.g.[26,27]. Additionally, these data biases may change over time. Although our database is based on a systematic and thorough search of all the data available to us from all regions covered, the data are still likely to be biased because there will have been intrinsic biases in the available data sources. For example, in this database, central India is under-represented in terms of recent research locales and it is hard to disentangle whether this is due to a lower number of ecologists focussing their studies there or if it is a justified skew as a result of biodiversity loss in this area. More recent records also show a bias toward threatened species and Protected Areas[15]. There are very few records of species absence although of course absence may be inferred if there are many records of other species in a particular locality. For a more detailed discussion of bias and missing data see Boakes *et al.*[15] and Boakes *et al.*[10].

Received: 21 January 2020; Accepted: 24 September 2020;
Published online: 13 October 2020

References

1. Convention on Biological Diversity. *COP 10 Decision X/2 - Strategic Plan for Biodiversity 2011–2020*. (Montreal, Canada, 2010).
2. Sheppard, C. The shifting baseline syndrome. *Mar. Poll. Bull.* **30**, 766–767 (1995).
3. Willis, K. J. *et al.* How can a knowledge of the past help to conserve the future? Biodiversity conservation and the relevance of long-term ecological studies. *Philos. T. R. Soc. B* **362**, 175–186 (2007).
4. Jetz, W., McPherson, J. M. & Guralnick, R. P. Integrating biodiversity distribution knowledge: toward a global map of life. *Trends Ecol. Evol.* **27**, 151–159 (2012).
5. Rondinini, C. *et al.* Global habitat suitability models of terrestrial mammals. *Philos. T. R. Soc. B* **366**, 2633–2641 (2011).
6. Hagemeijer, W. & Blair, M. *The EBCC Atlas of European Breeding Birds*. (T & AD Poyser, England, 1997).
7. Sullivan, B. L. *et al.* eBird: A citizen-based bird observation network in the biological sciences. *Biol. Conserv.* **142**, 2282–2292 (2009).
8. Collen, B. *et al.* Monitoring change in vertebrate abundance: the Living Planet Index. *Conserv. Biol.* **23**, 317–327 (2009).
9. Balmford, A. (Less and less) great expectations. *Oryx* **33**, 87–88 (1999).
10. Boakes, E. H., Fuller, R. A., McGowan, P. J. K. & Mace, G. M. Uncertainty in identifying local extinctions: the distribution of missing data and its effects on biodiversity measures. *Biol. Lett.* **12**, 20150824, https://doi.org/10.1098/rsbl.2015.0824 (2016).

11. McGowan, P. J. K. & Garson, P. J. The Galliformes are highly threatened: Should we care? *Oryx* **36**, 311–312 (2002).
12. Boakes, E. H., Fuller, R. A. & McGowan, P. J. K. The extirpation of species outside protected areas. *Conserv. Lett.* **12**, e12608 (2019).
13. Madge, S. & McGowan, P. *Pheasants, partridges and grouse: A guide to the pheasants, partridges, quails, grouse, guineafowl, buttonquails and sandgrouse of the world*. (Christopher Helm Publishers Ltd, 2002).
14. Boakes, E. H. *et al*. GalliForm: Galliformes occurrence records from the Indo-Malay and Palaearctic, 1800–2008. *The Global Biodiversity Information Facility* https://doi.org/10.15468/9825yw (2020).
15. Boakes, E. H. *et al*. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS Biol.* **8**, e1000385 (2010).
16. Gupta, G., Dunn, J., Sanderson, R., Fuller, R. & McGowan, P. J. K. A simple method for assessing the completeness of a geographic range size estimate. *Glob. Ecol. Conserv.* **21**, e00788 (2020).
17. Boakes, E. H., Isaac, N. J. B., Fuller, R. A., Mace, G. M. & McGowan, P. J. K. Examining the relationship between local extinction risk and position in range. *Conserv. Biol.* **32**, 229–239 (2018).
18. Mace, G. M., Collen, B., Fuller, R. A. & Boakes, E. H. Population and geographic range dynamics: implications for conservation planning. *Philos. T. R. Soc. B* **365**, 3743–3371 (2010).
19. Dunn, J. C., Buchanan, G. M., Cuthbert, R. J., Whittingham, M. J. & McGowan, P. J. K. Mapping the potential distribution of the Critically Endangered Himalayan Quail *Ophrysia superciliosa* using proxy species and species distribution modelling. *Bird Conserv. Int.* **25**, 466–478 (2015).
20. Dunn, J. C., Buchanan, G. M., Stein, R. W., Whittingham, M. J. & McGowan, P. J. K. Optimising different types of biodiversity coverage of protected areas with a case study using Himalayan Galliformes. *Biol. Conserv.* **196**, 22–30 (2016).
21. Grainger, M. J., Ngoprasert, D., McGowan, P. J. K. & Savini, T. Informing decisions on an extremely data poor species facing imminent extinction. *Oryx* **53**, 484–490 (2019).
22. Olson, D. M. *et al*. Terrestrial ecoregions of the world: A new map of life on earth. *Bioscience* **51**, 933–938 (2001).
23. Roselaar, C. S. An inventory of major European bird collections. *Bull. Br. Ornithol. Club* **123A**, 253–337 (2003).
24. McGowan, P. J. K. Mapping the distribution of Asian partridges and pheasants as a requirement for identifying conservation priorities. *Acta Zool. Sinica* **42S**, 11–19 (1996).
25. Wieczorek, J., Guo, Q. & Hijmans, R. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *Int. J. Geogr. Inf. Sci.* **18**, 745–767 (2004).
26. Isaac, N. J. B. & Pocock, M. J. O. Bias and information in biological records. *Biol. J. Linn. Soc.* **115**, 522–531 (2015).
27. Collen, B., Ram, M., Zamin, T. & McRae, L. The tropical biodiversity data gap: addressing disparity in global monitoring. *Trop. Conserv. Sci.* **1**, 75–88 (2008).
28. Barnes, R. dggridR: Discrete Global Grids for R. https://cran.r-project.org/web/packages/dggridR/ (2016).
29. R Core Team. R: Version 3.5.1. A Language and Environment for Statistical Computing. http://www.R-project.org (2018).
30. Sharrock, J.T.R. *The Atlas of Breeding Birds in Britain and Ireland* (Poyser, 1977).
31. Gibbons, D.W. *The New Atlas of Breeding Birds in Britain and Ireland* (Poyser, 1993).
32. Lord, J. & Munns, D.J. *Atlas of Breeding Birds of the West Midlands* (Collins, 1970).
33. Alamany, O., Auclair, R. & Bertrand, A. *Atlas of the Breeding Birds of Andorra* (Associasió per a la Defensa de la Natura, Andorra, 2002).
34. Devilliers, P. *et al*. *Atlas des Oiseaux Nicheurs de Belgique* (Institut Royale des Sciences Naturelles, 1988).
35. Melchior, E. *Atlas of Breeding Birds in Luxembourg* (Lëtzebuerger Natur- a Vulleschutzliga, 1987).
36. Teixeira, R.M. *Atlas van de Nederlandse Broedvogels 1973-1977* (Natuurmonumenten 1979).
37. Atlas van de Nederlandse Broedvogels 1978-1983 (Sovon, 1987).
38. Rufino, R. *Atlas das Aves que Nidificam em Portugal Continental* (ICNB/Assírio & Alvin 1989).
39. Florenzano, G.T., Arcamone, E., Baccetti, N., Meschini, E. & Sposimo, P. *Atlante Degli Uccelli Nidificanti e Svernanti in Toscana* (CentrOrnitologicoToscano, 1997).
40. Stastny, K., Randik, A. & Hudec, K. *Atlas Hnizdniho Rozsireni Ptaku V CSSR* (Academia, 1987).
41. Kalyakin, M.V. & Voltzit, O.V. *Birds of Moscow City and the Moscow Region* (Pensoft, 2006).
42. Renno, O. *Eesti Linnuatlas* (Light, 1993).
43. Priednieks, J., Strazds, M, Strazds, A. & Petrins, A. Latvian Breeding Bird Atlas (Riga Zinatne, 1989).
44. Sovinc, A. *Zimski Ornitoloski Atlas Slovenije* (Tehniška založba Slovenije, 1994).
45. Eriksen, J. *Breeding bird atlas of Oman* (Oman Bird Records Committee, 1998).
46. Jennings, M.C. *An Interim Atlas of the Breeding Birds of Arabia* (National Commission for Wildlife Conseravtion, 1995).
47. Nikolaus, G. *Distribution Atlas of Sudan's Birds with Notes on Habitat and Status* (Zoologisches Forschungsmuseum Alexander Koenig, 1987).
48. Atlas of Wintering Birds of Japan (Ministry of Environment, 2004).

Acknowledgements

This work was funded by grant number F/07/058/AK from the Leverhulme Trust. We are extremely grateful to the museums, libraries and ringing groups which shared their collections with us as well as the many taxonomic and regional experts who reviewed our data. We thank Hajir Al-Khairullah, Kate Harris, Cecilia Orme and Helen Pine who helped collect data but with whom we have since lost touch and could not offer co-authorship and also Pavel Tomkovich who facilitated data collection at the Zoological Museum, Lomonosov Moscow State University. Sophia Ratcliffe helped us convert the database to Darwin Core format.

Author contributions

E.H.B. wrote the manuscript, managed the data compilation, assisted with data collection and georeferencing and performed validation checks. R.A.F. shared previously compiled data, assisted with data collection and georeferencing and performed validation checks. G.M.M. oversaw and commented on the data compilation process. C.D. shared previously compiled data and assisted with data collection and georeferencing. T-T. A. constructed the database, coded the data-entry processes and assisted with data collection and georeferencing. A.A. assisted with data collection and georeferencing. N.E.C. assisted with data collection and georeferencing. J.D. assisted with data collection and georeferencing. G.G. assisted with data collection and georeferencing. J.G. assisted with data collection and georeferencing. V.G. assisted with data collection and georeferencing. U.I. assisted with data collection and georeferencing. E.J. assisted with data collection. K.O. assisted with data collection and georeferencing. E.P. assisted with data collection and georeferencing. R.P. assisted with data collection and georeferencing. J.S. assisted with data collection and georeferencing. S.S. assisted with data collection and georeferencing. T.T. assisted with data collection and georeferencing. P.J.K.M. conceived the idea, shared previously compiled data and performed validation checks.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to E.H.B.