

Data gaps and opportunities for comparative and conservation biology

Dalia A. Conde^{a,b,c,1}, Johanna Staerk^{a,b,c,d}, Fernando Colchero^{b,e}, Rita da Silva^{a,b,c}, Jonas Schöley^b, H. Maria Baden^{b,c}, Lionel Juvet^{b,c}, John E. Fa^f, Hassan Syed^g, Eelke Jongejans^h, Shai Meiriⁱ, Jean-Michel Gaillard^j, Scott Chamberlain^k, Jonathan Wilcken^l, Owen R. Jones^{b,c}, Johan P. Dahlgren^{b,c}, Ulrich K. Steiner^{b,c}, Lucie M. Bland^m, Ivan Gomez-Mestreⁿ, Jean-Dominique Lebreton^o, Jaime González Vargas^p, Nate Flesness^a, Vladimir Canudas-Romo^q, Roberto Salguero-Gómez^r, Onnie Byers^s, Thomas Bjørneboe Berg^t, Alexander Scheuerlein^d, Sébastien Devillard^j, Dmitry S. Schigel^u, Oliver A. Ryder^v, Hugh P. Possingham^w, Annette Baudisch^b, and James W. Vaupel^{b,d,x,1}

^aSpecies360 Conservation Science Alliance, Bloomington, MN 55425; ^bInterdisciplinary Center on Population Dynamics, University of Southern Denmark, 5230 Odense M, Denmark; ^cDepartment of Biology, University of Southern Denmark, 5230 Odense M, Denmark; ^dMax Planck Institute for Demographic Research, D-18057 Rostock, Germany; ^eDepartment of Mathematics and Computer Science, University of Southern Denmark, 5230 Odense M, Denmark; ^fDivision of Biology and Conservation Ecology, School of Science and the Environment, Manchester Metropolitan University, Manchester, M15 6BH, United Kingdom; ^gBir Ventures, Bloomington, MN 55425; ^hDepartment of Animal Ecology and Physiology, Radboud University, 6525 AJ Nijmegen, The Netherlands; ⁱDepartment of Zoology, Tel Aviv University, 69978 Tel Aviv, Israel; ^jDépartement de Génie Biologique, University of Lyon, 69622 Villeurbanne Cedex, France; ^kOpenSci, University of California Museum of Paleontology, Berkeley, CA 94720; ^lAuckland Zoo, Auckland 1022, New Zealand; ^mSchool of BioSciences, The University of Melbourne, Royal Parade, Parkville, VIC 3052, Australia; ⁿEstación Biológica de Doñana, Consejo Superior de Investigaciones Científicas, 41092 Sevilla, Spain; ^oCNRS, Centre d'écologie fonctionnelle et évolutive, UMR 5175 1919, 34293 Montpellier Cedex 5, France; ^pAbitar Learning Technologies, SC, Tlalpan, 14350 Mexico City, Mexico; ^qSchool of Demography, College of Arts and Social Sciences, Australian National University, Canberra, ACT 2600, Australia; ^rDepartment of Zoology, University of Oxford, OX2 6GG Oxford, United Kingdom; ^sConservation Breeding Specialist Group, Species Survival Commission, International Union for Conservation of Nature, Minneapolis, MN 55124; ^tNaturama, 5700 Svendborg, Denmark; ^uGlobal Biodiversity Information Facility, 2100 Copenhagen Ø, Denmark; ^vSan Diego Zoo Global Institute for Conservation Research, Escondido, CA 92027; ^wAustralian Research Council Centre of Excellence for Environmental Decisions, The University of Queensland, Brisbane, QLD 4072, Australia; and ^xDuke Population Research Institute, Duke University, Durham, NC 27705

Contributed by James W. Vaupel, March 12, 2019 (sent for review November 6, 2018; reviewed by Luigi Boitani and Deborah Roach)

Biodiversity loss is a major challenge. Over the past century, the average rate of vertebrate extinction has been about 100-fold higher than the estimated background rate and population declines continue to increase globally. Birth and death rates determine the pace of population increase or decline, thus driving the expansion or extinction of a species. Design of species conservation policies hence depends on demographic data (e.g., for extinction risk assessments or estimation of harvesting quotas). However, an overview of the accessible data, even for better known taxa, is lacking. Here, we present the Demographic Species Knowledge Index, which classifies the available information for 32,144 (97%) of extant described mammals, birds, reptiles, and amphibians. We show that only 1.3% of the tetrapod species have comprehensive information on birth and death rates. We found no demographic measures, not even crude ones such as maximum life span or typical litter/clutch size, for 65% of threatened tetrapods. More field studies are needed; however, some progress can be made by digitalizing existing knowledge, by imputing data from related species with similar life histories, and by using information from captive populations. We show that data from zoos and aquariums in the Species360 network can significantly improve knowledge for an almost eightfold gain. Assessing the landscape of limited demographic knowledge is essential to prioritize ways to fill data gaps. Such information is urgently needed to implement management strategies to conserve at-risk taxa and to discover new unifying concepts and evolutionary relationships across thousands of tetrapod species.

biodemography | mortality | fertility | extinction | Demographic Species Knowledge Index

Accessible data are increasingly becoming more valuable in research and for decision-making processes worldwide, including conservation. Most of the world's digitally available information has been compiled in the past few years, and data acquisition rates are accelerating (1). Collection and digitization of existing biodiversity data are essential for making more species information available to support conservation actions. Identifying knowledge gaps and catalyzing efforts to generate and use existing information have become priorities for international bodies concerned about the protection of global biodiversity

[e.g., the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (2)]. Furthermore, making these data available to scientists and practitioners is important

Significance

Given the current species extinction rates, evidence-based policies to conserve at-risk species are urgently needed. Ultimately, the extinction of a species is determined by birth and death rates, which drive populations to increase or decline. Therefore, demographic data are essential to inform species conservation policies or to develop extinction risk assessments. Demographic information provides an indispensable bedrock for insights to tackle species sustainable management and deepens understanding of ecological and evolutionary processes. We develop a Demographic Species Knowledge Index that classifies the demographic information for 32,144 tetrapod species. We found comprehensive information on birth and survival for only 1.3% (613) of the species, and show the major potential of zoos and aquariums to significantly increase our demographic knowledge.

Author contributions: D.A.C. and J.W.V. designed research; D.A.C., J. Staerk, F.C., R.d.S., and H.S. performed research; D.A.C., J. Schöley, L.J., H.S., E.J., S.M., S.C., O.R.J., J.P.D., U.K.S., L.M.B., I.G.-M., J.-D.L., J.G.V., N.F., V.C.-R., R.S.-G., O.B., T.B.B., A.S., S.D., D.S.S., O.A.R., H.P.P., A.B., and J.W.V. contributed new reagents/analytic tools; D.A.C., J. Staerk, F.C., R.d.S., H.S., and J.W.V. analyzed data; D.A.C., J. Staerk, F.C., J. Schöley, H.M.B., J.E.F., J.-M.G., D.S.S., and J.W.V. wrote the paper with contribution from all authors for final version; S.M., O.R.J., J.P.D., I.G.-M., J.-D.L., A.S., and S.D. contributed datasets to the paper; T.B.B. and D.S. contributed ideas on the use of museums.

Reviewers: L.B., Università di Roma Sapienza; and D.R., University of Virginia.

The authors declare no conflict of interest.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Data deposition: Data to perform the analyses have been deposited in the Species360 Open Data Portal with additional figures (<https://www.species360.org/serving-conservation/species-knowledge-index/>) and in the Dryad Digital Repository (<https://doi.org/10.5061/dryad.nq02fm3>).

¹To whom correspondence may be addressed. Email: dalia.conde@species360.org or jvaupel@sdu.dk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1816367116/-DCSupplemental.

Published online April 19, 2019.

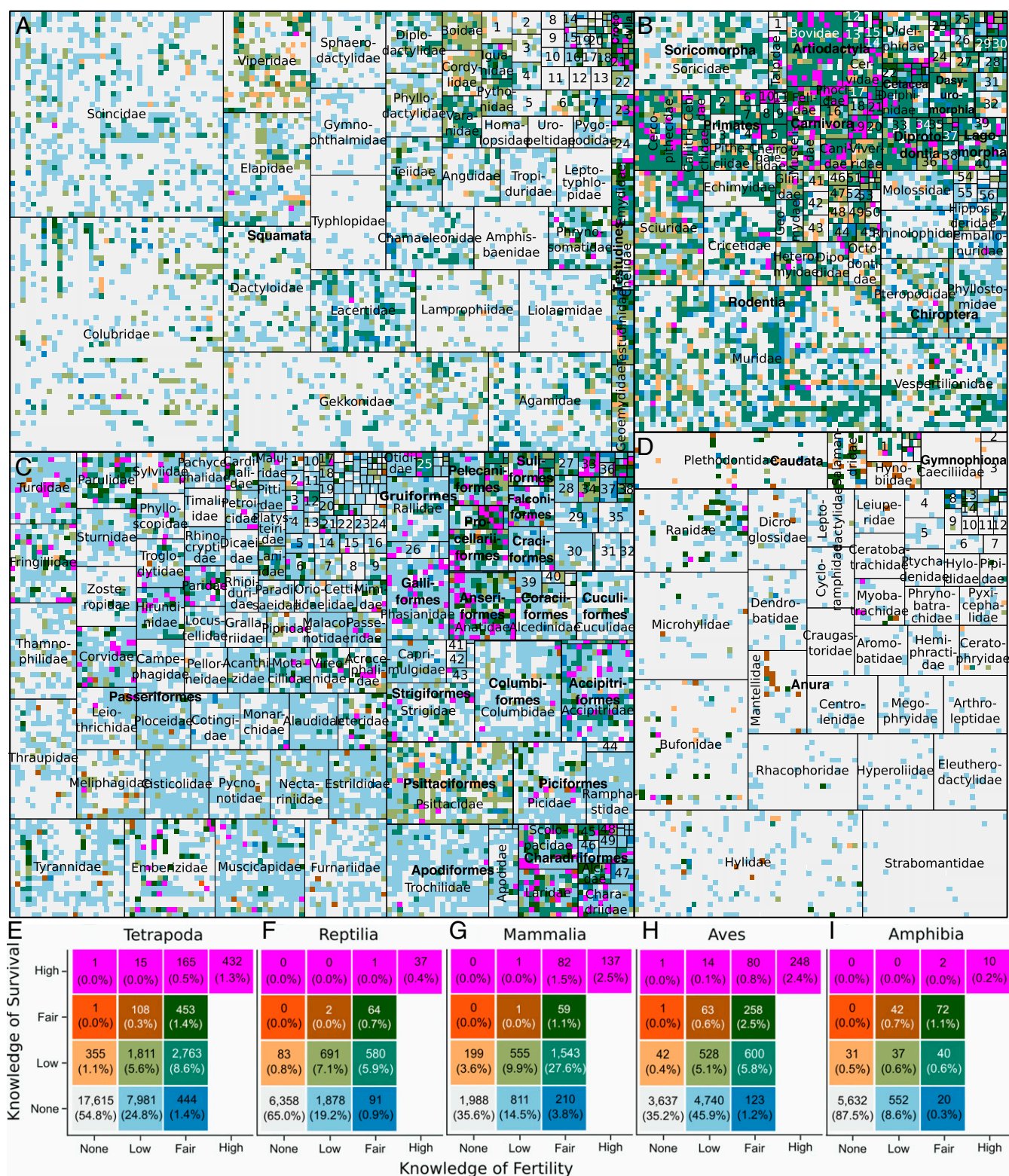


Fig. 1. Landscape of demographic knowledge for tetrapods. (A) Reptilia. (B) Mammalia. (C) Aves. (D) Amphibia. Each pixel represents a species, hierarchically ordered by families, orders, and classes. The level of information on fertility and survival is coded using a 2D color scale, with blue shades representing information on fertility and red shades representing information on survival. Green shades represent equal information on both. When only one measure was available, knowledge was classified as low. When two or more measures were available, knowledge was classified as fair. Knowledge was classified as high when detailed age-specific or stage-specific information was available in a life table or population matrix, indicated by the pink shade. Gray indicates no information. Squares show the number of species and percentages per index for all tetrapods (E) and divided by class (F–I).

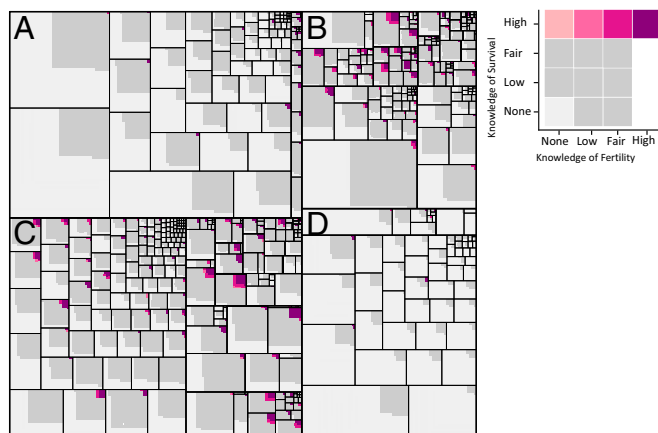


Fig. 2. Simplified version of the landscape shown in Fig. 1. (A) Reptilia. (B) Mammalia. (C) Aves. (D) Amphibia. Pink shades represent high knowledge of survival and various levels of knowledge about fertility. Dark gray shades represent low or fair knowledge, and the light gray areas indicate no demographic knowledge. For the entire range of tetrapods, only 1.3% of species have high survival and fertility information, less than 0.6% have high survival but little or no fertility information, 43.3% have limited survival and fertility information, and 54.8% have no survival or fertility information.

for Conservation of Nature Red List of Threatened Species (hereafter IUCN Red List) (11), which is the average age of mothers at the birth of offspring, and which provides a measure of the time required for a population to renew itself. Estimation of generation length ideally requires knowledge of age- or stage-specific survival and fertility. Likewise, to set up harvesting quotas, it is necessary to predict the impact of harvesting on the sustainability of a population. Therefore, population viability analyses are often required; these preferably use detailed measures of age- or stage-specific survival and fertility because these measures greatly improve estimation of population trends under different management scenarios and the prediction of extinction risk (12). For example, CITES, the Convention on International Trade in Endangered Species of Flora and Fauna usually requires these types of analyses for the establishment of exporting quotas for particular species to ensure that the international trade does not threaten the sustainability of their populations.

Detailed demographic data are essential not only for managing populations but also for understanding life histories and population dynamics. For example, age-specific mortality and fertility data are crucial for studies of the biology of aging in humans and nonhuman species (6, 7). Moreover, the patchy nature of the landscape of demographic knowledge is especially worrisome for threatened species for which data on closely related species are also lacking, as clearly illustrated by amphibians. After surviving four mass extinctions, amphibians now suffer the highest disappearance rate of all tetrapod classes (13). It is important that data gaps like these are filled by collection of field data, when possible; otherwise, data from captive populations can provide important information or estimates can be derived from closely related species.

Imputation methods are often used to fill information gaps when data on related species are available. These methods estimate missing data by using suites of trait correlations among species (14). For example, if detailed demographic measures are not available, simple life history traits, such as body size, have been used to make crude predictions of extinction risk. For highly data-deficient groups, a potential source of information lies in the availability of demographic and related measures from natural history museum collections, such as number of embryos in the uterus from preserved specimens, age estimates based on the characteristics of skulls or teeth, skeletal indicators of health,

and size and weight of individuals at the time of capture. The incorporation of existing demographic data from unpublished studies, reports, and journals in languages other than English, as well as data from captive populations, will also play a key role in filling knowledge gaps.

To inform animal management decisions, zoos and aquariums collect detailed information on individuals under their care. For 45 y, Species360 has been gathering standardized information from institutions worldwide; currently, information is available for over 10 million individuals from 22,000 species (15). We found that the use of Species360 members' data could significantly increase knowledge, such as age at first reproduction from 4,199 species to 7,273 species, a 73% increase. More dramatically, the availability of life tables or population matrices could be increased from 613 species to 4,699 species, an almost eightfold gain.

Caution must be taken when using data from captive populations to model wild populations. Zoo and aquarium populations are intensively managed, and hence likely to differ from free-living populations, notably in survival (16) and reproduction metrics. Furthermore, we found that origin of the information for more than half of the species (66%) is unknown or not reported (Fig. 3 and *SI Appendix*, Table S3). Therefore, whether demographic measures were estimated from imputation analyses or from wild or captive populations is unclear (Fig. 3 and *SI Appendix*, Fig. S5). We found that between 75% and 85% of the species have an unknown or not reported origin of information for interlitter or interbirth interval, age at first reproduction, and litter or clutch size (Table 4). Likewise, 57% of the species have an unknown origin for maximum life span. This is worrisome because these data are widely used for conservation and comparative studies. Thus, gaining a better understanding of biases of data from unknown origin, imputation analyses, or populations under captive management should be a priority. In addition, it will be important to explore the uncertainty introduced by mixing data from wild and captive populations. In this sense, zoos, aquariums, and botanical gardens could become key allies in providing data that can help fill data gaps to understand species biology.

To address current biodiversity crises, key questions must be answered. Which species should be selected for long-term population

Table 2. Number of species per Demographic Species Knowledge Index and IUCN Red List categories

Demographic Species Knowledge Index	Survival Fertility	IUCN Red List category									
		LC	NT	VU	EN	CR	EW	EX	DD	NE	Total
None	None	6,609	977	1,220	1,331	771	5	132	2,484	4,086	17,615
None	Low	5,306	394	363	278	107	2	14	146	1,371	7,981
None	Fair	274	33	26	37	14	0	1	9	50	444
Low	None	169	20	39	19	13	1	2	26	66	355
Low	Low	1,031	105	117	89	37	0	8	51	373	1,811
Low	Fair	1,601	166	235	179	82	3	14	75	408	2,763
Fair	None	0	0	0	0	0	0	0	0	1	1
Fair	Low	69	8	9	7	2	0	0	2	11	108
Fair	Fair	305	31	31	20	8	0	0	0	58	453
High	None	1	0	0	0	0	0	0	0	0	1
High	Low	9	0	2	1	0	0	0	0	3	15
High	Fair	121	8	12	7	3	0	0	0	14	165
High	High	281	34	39	23	15	0	0	1	39	432
Total		15,776	1,776	2,093	1,991	1,052	11	171	2,794	6,480	32,144

CR, critically endangered; DD, data deficient; EN, endangered; EW, extinct in the wild; EX, extinct; IUCN, International Union for Conservation of Nature; LC, least concern; NE, not evaluated; NT, near threatened; VU, vulnerable. Further information about measures of knowledge for the Demographic Species Knowledge Index categories is provided in *Methods*.

Table 4. Number of species indicating the origin (captive, wild, or/and unknown) from which demographic measures or rates were estimated for all tetrapods

Demographic measure	Origin/no. of species		
Fertility	Wild	Captive	Unknown
Age at first reproduction	1,222	43	4,114
Interlitter/interbirth interval	402	10	1,256
Litter/clutch size	2,413	31	13,735
Proportion of reproductive females	44	0	0
Recruitment	22	0	0
Age- or stage-specific fertility rates	416	14	22
Survival			
Maximum recorded life span	1,483	2,358	5,128
Mean age of (adult) population	0	0	114
Crude mortality	1,055	13	229
Age- or stage-specific death rate	580	48	26

A single species may have data from different origins.

accepted names and the IUCN Red List status (23), we used the taxize (24) package in R version 3.5.1 (25) and manually searched for species names that could not be retrieved. For 3% of the species, we were not able to resolve their taxonomy, so they were not included in the analyses. This process resulted in a metadatabase of 32,144 species, with 14,529 species with demographic data and 115,356 demographic records. We standardized each record's origin from populations reported as wild, captive, or unknown across all of the databases. When the origin was not provided in the database, we assigned it as "unknown" (Table 4); however, we still included those records because all of the databases included here have a reference to a publication.

Developing the Demographic Species Knowledge Index. To summarize the availability of demographic data for each tetrapod species we developed the Demographic Species Knowledge Index. This index provides scores that summarize the number of a total of eight measures available for fertility and survival for any species. These measures are as follows:

- Measures of fertility knowledge: (i) age at first reproduction; (ii) interlitter/interbirth interval; (iii) litter/clutch size; (iv) proportion of adult females that are reproductive; and (v) birth or recruitment rate, with recruitment denoting the average number of individuals that reach a specific age or stage (e.g., maturity, leaving the nest) per reproductive female.
- Measures of survival knowledge: (i) maximum recorded life span, (ii) mean age of the (adult) population, and (iii) crude mortality. Information about mortality (or survival) includes the juvenile crude death rate, the adult crude

death rate (or adult life expectancy, approximately the inverse of the adult crude death rate), and the crude death rate for juveniles and adults combined (or life expectancy at birth, which is approximately its inverse). The crude death rate is given by the number of deaths in some time interval over average population size in the interval. The probability of death equals the number of deaths in some time interval divided by population size at the beginning of the interval. Biologists sometimes refer to one minus either of these measures as the survival rate.

- Combined age or stage survival-fertility knowledge: The index is also based on the availability of population-level data in the form of population matrices or life tables. These include both age- or stage-specific death or survival probabilities and age- or stage-specific fertility rates; life tables often contain only mortality data.

Knowledge about survival is classified into four categories:

- High: A life table or population matrix is available.
- Fair: Such data are not available, but at least two variables are measured, such as maximum life span and adult mortality.
- Low: Only one variable is available.
- None: No information is available.

Knowledge about fertility is also classified into four categories.

- High: Fertility rates are available by age or stage.
- Fair: Such data are not available, but at least two variables are measured, such as age at maturity and average litter/clutch size.
- Low: Only one variable is available.
- None: No information is available.

Life tables and matrices always contain survival information but do not always have information on fertility, which is usually harder to obtain in the wild. Hence, in Fig. 1, only 13 categories are color-coded. The metadatabase to estimate the index and the index are both available in the Species360 Open Data Portal and Dryad Digital Repository (48, 49).

ACKNOWLEDGMENTS. We especially thank Bengt Holst for facilitating the expert workshop for standardization of the demographic vocabulary across the 22 data repositories used here. We thank Jon Paul Rodriguez, Yvonne Buckley, Jim Guenter, Iain Stott, Simeon Q. Smelee, Stephan M. Funk, and the two reviewers for useful comments on the manuscript. We thank the commitment of zoo and aquarium staff for managing their animal records in Zoological Information Management System (ZIMS), providing vast amounts of data to create a wealth of demographic knowledge. We acknowledge the following institutions for their financial support: Interdisciplinary Centre on Population Dynamics, Max Planck Institute for Demographic Research, Bir Ventures, University of Southern Denmark, Gerhard und Ellen Zeidler-Stiftung, and Sponsor-Partners of the Species360 Conservation Science Alliance (the Copenhagen Zoo, World Association of Zoos and Aquariums, and Wildlife Reserves of Singapore).

- Sagiroglu S, Sinanc D (2013) Big data: A review. *International Conference on Collaboration Technologies and Systems (CTS)* (IEEE, Piscataway, NJ), pp 42–47.
- Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (2016) Knowledge, Information and Data. Available at <https://www.ipbes.net>. Accessed April 3, 2019.
- Convention on Biological Diversity (2016) Aichi Biodiversity Targets, target 19. Available at <https://www.cbd.int/aichi-targets/target/19>. Accessed April 3, 2019.
- Stephenson PJ, et al. (2017) Priorities for big biodiversity data. *Front Ecol Environ* 15:124–125.
- Carey JR, Judge DS (2000) *Life Spans of Mammals, Birds, Amphibians, Reptiles, and Fish*. Odense Monographs of Population Aging 8. (Odense Univ Press, Odense, Denmark). Available at https://www.demogr.mpg.de/en/projects_publications/publications_1904/monographs/life_spans_of_mammals_birds_amphibians_reptiles_and_fish_1005.htm. Accessed April 3, 2019.
- Colchero F, et al. (2016) The emergence of longevous populations. *Proc Natl Acad Sci USA* 113:E7681–E7690.
- Jones OR, et al. (2014) Diversity of ageing across the tree of life. *Nature* 505:169–173.
- Park CA, et al. (2013) The vertebrate trait ontology: A controlled vocabulary for the annotation of trait data across species. *J Biomed Semantics* 4:13.
- Roskov Y, et al. (2018) *Species 2000 & ITIS Catalogue of Life, 2018 Annual Checklist* (Species 2000: Naturalis, Leiden, The Netherlands). Available at www.catalogueoflife.org/annual-checklist/2018. Accessed January 30, 2018.
- Ceballos G, Ehrlich PR, Dirzo R (2017) Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proc Natl Acad Sci USA* 114:E6089–E6096.
- Staerk J, et al. (2019) Performance of generation time approximations for extinction risk assessments. *J Appl Ecol*, 10.1111/1365-2664.13368.
- Colchero F, et al. (2019) The diversity of population responses to environmental change. *Ecol Lett* 22:342–353.

- Wake DB, Vredenburg VT (2008) Colloquium paper: Are we in the midst of the sixth mass extinction? A view from the world of amphibians. *Proc Natl Acad Sci USA* 105: 11466–11473.
- Hilbers JP, et al. (2016) An allometric approach to quantify the extinction vulnerability of birds and mammals. *Ecology* 97:615–626.
- Species360 (2016) Global information serving conservation. Available at <https://www.species360.org>. Accessed December 1, 2018.
- Tidière M, et al. (2016) Comparative analyses of longevity and senescence reveal variable survival benefits of living in zoos across mammals. *Sci Rep* 6:36361.
- Knight AT, et al. (2010) Barometer of life: More action, not more data. *Science* 329: 141, author reply 141–142.
- Wieczorek J, et al. (2012) Darwin Core: An evolving community-developed biodiversity data standard. *PLoS One* 7:e29715.
- Addink W, Koureas D, Casino A (2018) DiSSCo: The physical and data infrastructure for Europe's natural science collections. *European Geosciences Union General Assembly Conference Abstracts* 20:16356.
- Legalatlas (2018) Understand the law. Then use it. Available at <https://www.legalatlas.net/>. Accessed April 3, 2019.
- Strier KB, et al. (2010) The primate life history database: A unique shared ecological data resource. *Methods Ecol Evol* 1:199–211.
- Parr CS, et al. (2016) TraitBank: Practical semantics for organism attribute data. *Semant Web* 7:577–588.
- International Union for Conservation of Nature (2018) The IUCN Red List of Threatened Species. Version 2018-2. Available at <https://www.iucnredlist.org>. Accessed March 27, 2019.
- Chamberlain SA, Szöcs E (2013) Taxize: Taxonomic search and retrieval in R. *F1000 Res* 2:191.
- R Core Team (2018) R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna).

