POLICY FORUM

NATURAL HISTORY

A global approach for natural history museum collections

Integration of the world's natural history collections can provide a resource for decision-makers

By **Kirk R. Johnson[1], Ian F. P. Owens[2], the Global Collection Group**

Over the past three centuries, people have collected objects and specimens and placed them in natural history museums throughout the world. Taken as a whole, this global collection is the physical basis for our understanding of the natural world and our place in it, an unparalleled source of information that is directly relevant to issues as diverse as wildlife conservation, climate change, pandemic preparedness, food security, invasive species, rare minerals, and the bioeconomy (*1*). Strategic coordination and use of the global collection has the potential to focus future collecting and guide decisions that are relevant to the future of humanity and biodiversity. To begin to map the aggregate holdings of the global collection, we describe here a simple and fast method to assess the contents of any natural history museum, and report results based on our assessment of 73 of the world's largest natural history museums and herbaria from 28 countries.

Today, more than a thousand natural history museums exist, with the largest ones located in Europe and North America. The world's natural history collections provide a window into the planet's past and present, and they are increasingly being used to make actionable predictions relative to climate change, biodiversity loss, and infectious disease. For example, natural history museum data are the fundamental source of primary biodiversity knowledge underlying major policy frameworks. The 2018 Intergovernmental Panel on Climate Change (IPCC) Special Report on Global Warming of 1.5°C used over 385 million species occurrence records, aggregated and tracked by the Global Biodiversity Information Facility (GBIF), from 5432 data providers, mostly natural history museums (*2, 3*), to show species movement in response to climate change [see supplementary materials (SM) for additional case studies].

[1]National Museum of Natural History (NMNH), Smithsonian Institution, Washington, DC, USA. [2]Cornell Laboratory of Ornithology, Cornell University, Ithaca, NY, USA. The Global Collection Group members and affiliations are listed in the supplementary materials. Email: johnsonkr@si.edu

Yet despite their enormous potential value to society, the information embedded in the collections housed in these museums is largely inaccessible. Fortunately, advances in digital, isotopic, imaging, and genomic technologies, as well as machine learning and artificial intelligence, are transforming and amplifying how natural history collections can be accessed and used (*1*). These innovations are substantially broadening the range of possible applications to include human health, cultural revitalization, and environmental monitoring. Increasingly, Indigenous interlocuters are joining these conversations and enriching them (*4, 5*).

In the past few decades, several networks have increased cooperation between biodiversity-based institutions around the world. In addition to GBIF, the Taxonomic Databases Working Group (TDWG), the Global Genome Biodiversity Network (GGBN), the Catalogue of Life (COL), the Earth BioGenome Project (EBP), the International Barcode of Life (iBOL), and the Biodiversity Heritage Library (BHL) have provided global leadership for integrating specimen data, taxonomic observations, genomes, and published literature on the natural world. Guiding principles for governing such data have emerged for traditional [e.g., FAIR (*6*)] and nontraditional users [e.g., CARE (*4*)]. Atlas of Living Australia (ALA) and Integrated Digitized Biocollections (iDigBio) in the United States represent successful national programs that develop innovative solutions to support collection digitization, data integration, and mobilization. They have fostered integration among stakeholders by making large datasets readily accessible. Other successful initiatives include the South African National Biodiversity Institute (SANBI) network, speciesLink (CRIA) in Brazil, and the National Commission for the Knowledge and Use of Biodiversity (CONABIO) in Mexico.

Although these institutions and efforts are playing vital roles in aggregating data, they do not create the collections and fill gaps therein. It is the natural history museums that actively curate and expand the collections. Thus, it falls on the museums to lead the way to deploy strategic collecting in service of future collection and policy outcomes. It will not be possible to do this unless museums understand the present scope of the global collection and thus its gaps. Yet natural history museums have generally operated independently, and no interoperable data structure exists to provide open access to their collective holdings. Because most natural history museum data are not digitally discoverable, the networks of data aggregators have not been able to access these "dark data" (*7*), the majority of museum specimens and objects that are the physical basis of natural history and cultural knowledge.

As the first step toward building a global network, we worked with the directors and lead science and collection staff of 73 of the world's largest natural history museums and herbaria from 28 countries to design and complete a simple and rapid survey of their collective holdings (see the first figure). Until now, it has been difficult to enumerate or compare the complete contents of large museums because their collections are not fully digitized, and the terminology used to describe subcollections is variable. Each of the 73 museums did report a total specimen count, and the sum of these counts was 1,147,934,687. We then subdivided this aggregate collection by creating a shared vocabulary for collection types and their geographic source areas. The result is a grid of 19 collection types by 16 geographic regions, such that any collection object from anywhere in the world would fall into only one of the resulting 304 cells (e.g., African insects; see the second figure). For this effort, the term "collection unit" represents a single museum's holdings within a single cell.

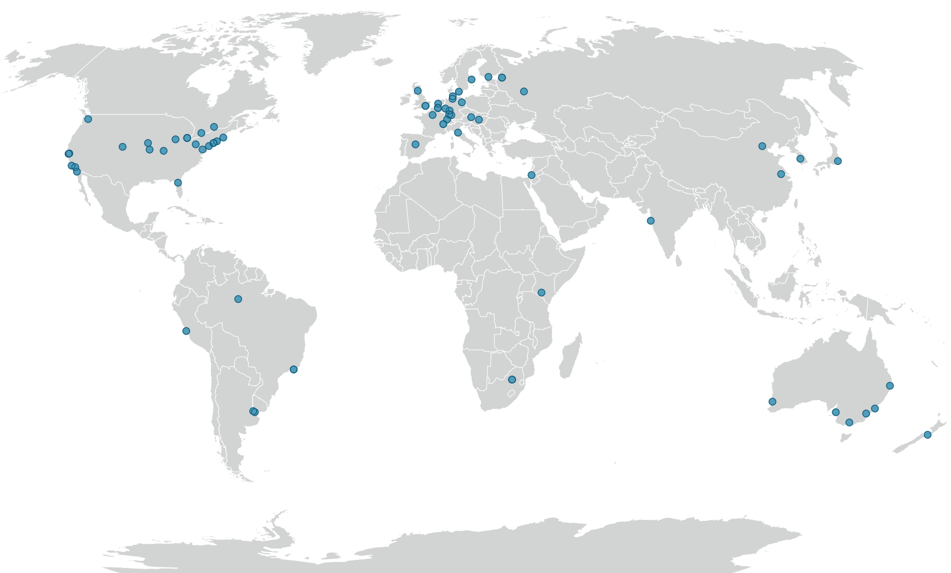
We then worked with expert staff from each museum to estimate the number of objects in each collection unit to the nearest order of magnitude (see the second figure). Because it is based on curatorial knowledge of each collection rather than catalog records, this approach is very rapid (<2 weeks for most museums). The value of this coarse-grained approach is that it allows museums to identify their largest collection units and define the strengths of their collections relative to those at other museums.

Heat mapping of collection units demonstrates the aggregated effort of the sampled museums and highlights regional and taxonomic focal areas and gaps (see the second figure). Most of the collection information that we surveyed is not digitally accessible: Only 16% of the objects have digitally discoverable records, and only 0.2% of biological collections have accessible genomic records.

We also surveyed the size and age distribution of the museum workforce that studies and cares for collections and that makes them available to the global community of users and found that the collections at the 73 museums and herbaria were tended by over 4500 science staff and nearly 4000 volunteers. See SM for further details on methods and data.

Locations of participating natural history museums

These 73 museums (●) collectively hold more than 1.1 billion objects.



GRAPHIC: D. AN-PHAM/*SCIENCE*

ACTIVATING THE GLOBAL COLLECTION

Our assessment allowed us to begin to map the aggregate holdings of the global collection, including the source areas and present locations of >1 billion objects (see the second figure and figs. S1 and S2). At the same time, it revealed many gaps, challenges, and opportunities. Work now needs to happen at a pace and magnitude that will meet the urgency of the Anthropocene and with the understanding that there are more species at risk of extinction than are currently known to science. Yet despite their potential value, natural history collections are at risk. Fires, natural disasters, and human conflicts can damage and destroy collections. Less pronounced degradation and destruction occur because of long-term underinvestment in infrastructure and expertise (*1*). We must invest in protecting and preserving these collections, and in expanding and integrating them, and associated expertise, with focused collecting efforts and new technologies such as genomics, environmental DNA, and artificial intelligence.

The concentration of the largest collections in European and North American cities remains a major barrier to connecting much of the world with its natural heritage, exacerbating the societal wrongs of colonial history in museum science. The same is true for museum expertise, where there is underinvestment relative to the potential of the collections and opportunities for training (fig. S3). Such support and coordination will be essential to overcome imbalances in information, expertise, and cultural differences.

Looking forward, we make the following recommendations [in line with those articulated in (*1, 8, 9*)] as a roadmap for museums, funders, governments, and industry to accelerate and coordinate their efforts while there is a window of opportunity in the next few decades to alter the trajectories of climate change and biodiversity loss.

Collecting in a rapidly changing planet

In the year 2100, scientists and policy-makers will look back to the collections made in the 21st century to inform their decisions about the 22nd century. Natural history museums must thus focus on the future. For example, the recent Global Biodiversity Framework (*10*) bridges the Convention on Biological Diversity, the Intergovernmental Panel on Biodiversity and Ecosystem Services, and other efforts by goal-setting 2030 as the time frame for changing trends in biodiversity loss. This is an opportunity to integrate museum datasets with conservation efforts. Yet even now, many regions of the world and many biological groups remain insufficiently studied and documented. Further analysis of the global collection will refine our understanding of the gaps and provide a roadmap for future collecting efforts. Toward that end, it is our intention to work with GBIF to make our data available to all natural history museums by integrating it with the Global Registry of Scientific Collections (GRSciColl).

For example, although our data show that the scale of the global collection is vast, it also highlights conspicuous gaps with respect to tropical and polar regions, marine systems, and undiscovered arthropod and microbial diversity. Also, our study does not address the hundreds of smaller museums, their collections, and their staffs, which comprise the rest of the global collection; these are especially valuable because of their regional holdings and the specificity of their expertise (*8*). Further, many current data aggregation efforts have largely focused on biodiversity and have been less inclusive of data from Earth and planetary sciences and anthropology—two fields that are traditionally major components of natural history collections.

New roles for old institutions

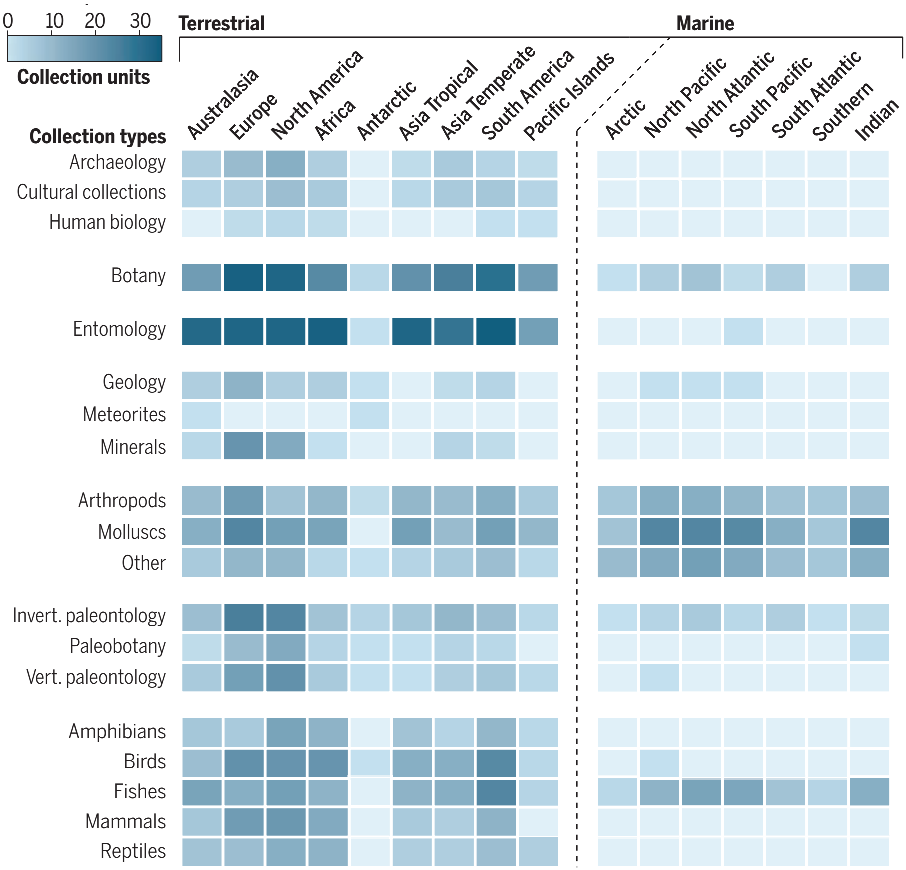
The collections that natural history museums acquired over the past three centuries carry data that cannot be replicated but can be used as baselines for efforts to regenerate ecosystems and revitalize communities. To facilitate efforts to preserve, protect, and expand collections, natural history museums must coordinate with initiatives such as GBIF, ALA, iDigBio, Distributed System of Scientific Collections (DiSSCo), and Biodiversity Collections Network (BCoN) to accelerate data practices, data sharing, and data stewardship in common knowledge platforms that provide an open informatics infrastructure to build a network of extended digital specimens (*9*). This step will happen through (i) digitization, which is ongoing and should be accelerated through high-throughput pipelines (e.g., conveyor belts capable of imaging >6000 herbarium sheets per day) and by prioritizing collections that have the greatest potential to mitigate global crises; (ii) genomic collections (including collections of frozen gametes and stem cells), which can be coordinated with the EBP and GGBN to facilitate sequencing of genomes across the tree of life; (iii) informatics, which can illuminate and interlink collections, making their information available for societal benefit; and (iv) training and capacity building, which can deliver reciprocal benefits, expertise, and the democratization of access to technologies.

Evolving ethics

Many historical practices used to acquire collections do not meet today's values and

Heatmap of global collection units

Any collection object in any museum can be categorized into only one of the 304 cells (19 collection types by 16 geographic regions). A "collection unit" is a single museum's holdings within a single cell. For 73 museums, there are 22,192 possible collection units. The heatmap shows the 1957 collection units with more than 10,000 objects. See supplementary materials for details and for a heatmap of the 242 collection units with more than 1 million objects.



GRAPHIC: D. AN-PHAM/*SCIENCE*

ethical standards for social and environmental justice. Many collections hold objects that may have been collected without prior consent, without permission from source communities, or without full recognition of contributors and cultural sensitivities (*4, 5, 11*). To redress past injustices, museums must engage with source communities to understand their perspectives and needs regarding objects and their associated metadata that may include sensitive cultural or ecological knowledge (*4, 12*). Ethical collecting in the 21st century requires local community engagement and benefit sharing that will also inspire and train the next generation of global scientists. There has been progress in this arena through national and international return efforts, which are largely focused on human remains and cultural artifacts (*13*), and there is increasing willingness to scrutinize the provenance of nonanthropological collections. The concept of shared stewardship is emerging as an opportunity to forge new partnerships around collection items where there are opportunities for mutual benefit. Growth and use of museum collections will be delayed or derailed unless there is real progress in this realm.

INTEGRATING INFRASTRUCTURE

The natural history museums of the world all hold portions of the global collection, a fact that argues for a new age of networked museums whose goal is to exceed the sum of their parts by strategic use of existing collections and thoughtful coordination of future collecting efforts. Yet despite several country-level initiatives to digitize specimens, most museum specimen information remains dark data (*7*). There is no single shared portal covering the breadth of life, Earth, and anthropological specimens in natural history collections, nor a way for researchers to link these data with other sources of information. We envision a coordinated strategy for the global collection that is based on strategic collecting, increased digitization, new technologies, and enhanced networking and coordination of museums. This strategy must embrace smaller regional collections, which bridge global gaps and provide critical insights and local context with source communities and geographies. Integrating this global infrastructure will also improve long-standing inequities in training and career opportunities and elevate marginalized voices.

Natural history collections are a form of science infrastructure that is necessary to support society-wide solutions. We hope to catalyze collaborative collection-based efforts to support future global sustainability, biodiversity, and climate frameworks (*10*, *11*) and identify critical gaps that might preclude such impact. Such collection-based efforts are typically not high-level institutional priorities; thus, the involvement of museum leadership, such as in our effort, is critical, particularly for mobilizing resources and coordinating efforts to strategically fill gaps in collections. ■

REFERENCES AND NOTES

1. Committee on Biological Collections, Their Past, Present, and Future Contributions and Options for Sustaining Them, Board on Life Sciences, Division on Earth and Life Studies, National Academies of Sciences, Engineering, and Medicine, *Biological Collections: Ensuring Critical Research and Education for the 21st Century* (National Academies Press, 2020); https://www.nap.edu/catalog/25592.
2. R. Warren, J. Price, E. Graham, N. Forstenhaeusler, J. VanDerWal, *Science* **360**, 791 (2018).
3. IPCC, "Summary for Policymakers. Global Warming of 1.5°C: IPCC Special Report on impacts of global warming of 1.5°C above pre-industrial levels in context of strengthening response to climate change, sustainable development, and efforts to eradicate poverty" (Cambridge Univ. Press, 2018).
4. S. R. Carroll *et al.*, *Data Sci. J.* **19**, 43 (2020).
5. L. Liggins, M. Hudson, J. Anderson, *Mol. Ecol.* **30**, 2477 (2021).
6. M. D. Wilkinson *et al.*, *Sci. Data* **3**, 160018 (2016).
7. C. R. Marshall *et al.*, *Biol. Lett.* **14**, 20180431 (2018).
8. A. K. Monfils *et al.*, *Bioscience* **70**, 1045 (2020).
9. M. Webster, J. Buschbom, A. Hardisty, A. Bentley, *Biodivers. Inf. Sci. Stand.* **5**, e75736 (2021).
10. UNEP Press Release, "COP15: Nations Adopt Four Goals, 23 Targets for 2030 in Landmark UN Biodiversity Agreement"; https://www.cbd.int/article/cop15-cbd-press-release-final-19dec2022.
11. Secretariat of the Convention on Biological Diversity United Nations Environmental Programme, *Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity* (2011).
12. M. Hudson *et al.*, *Nat. Rev. Genet.* **21**, 377 (2020).
13. L. Wade, *Science* **373**, 148 (2021).

ACKNOWLEDGMENTS

The authors thank the collections staff from participating institutions for providing data from their collections. We thank GBIF, GGBN, Consortium of European Taxonomic Facilities (CETAF), and iDigBio for assistance with access to information. We thank C. Butler, R. Corrigan N. Drew, H. Little, C. A. Meyer, and R. Snyder at the NMNH and staff at the Natural History Museum (NHM), London, for their technical assistance with data collation and C. Baldwin, J. Bell, C. Dove, S. Miller, E. Stofan, I. Krupnik, and E. Strong for comments and advice. We thank G. Mérite for information design and data illustration of the figures. Funding was provided by NMNH, NHM London, and the American Museum of Natural History. Data for the global collections survey have been published under an open license (Creative Commons Attribution 4.0 International), and can be persistently accessed at https://doi.org/10.5281/zenodo.6985399. Any personally identifiable information was redacted from the dataset prior to publication. A visualization dashboard of these data can be accessed at https://rebrand.ly/global-collections.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf6434

10.1126/science.adf6434