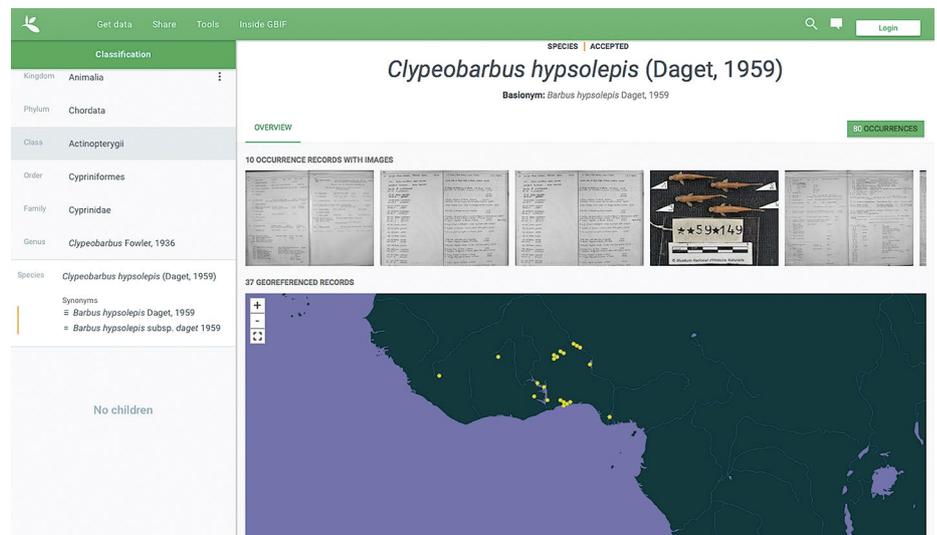


# Digitization of museum collections holds the potential to enhance researcher diversity

**To the Editor** — Natural history collections preserve critical information on biodiversity in space and time, but collections housed in museums have traditionally been inaccessible to many. At the same time, there remains great under-representation of many groups among researchers in science and education, even though diverse research teams approach problems with greater combinations of expertise and background, often resulting in more highly cited papers<sup>1</sup>. Here, we highlight how the advent of digitization (open access to images and specimen data) now makes a wealth of biodiversity information broadly available, and represents one method that can simultaneously increase access to those samples and help diversify our community. Digitization allows access to museum holdings to those for whom collections have typically been out of reach.

There is an unequal distribution of large scientific collections, with most of the major holdings being held in the Northern Hemisphere. Many of these collections in museums of the Global North were built from mega-diverse countries in the Global South through an integrated legacy of scientific exploration and colonialism. In some cases, the ecosystems of the past have been so altered that museum holdings are their only record<sup>2</sup>. Digitizing collections is one way in which their unequal distribution can potentially be mitigated, by making samples collected in biodiverse countries accessible to researchers and policymakers from those countries<sup>3</sup>. This access, therefore, can help to develop in-country taxonomic and systematic research programmes, a key component in quantifying global biodiversity loss<sup>4</sup>.

Digitization of collections may also improve researcher diversity through an additional avenue. Bringing multiple collections together through bioinformatic portals such as the Global Biodiversity Information Facility (GBIF; [www.gbif.org](http://www.gbif.org)) allows researchers from across the globe to simultaneously access many collections and synthesize large-scale, cross-taxa, biogeographic patterns, without requiring funding to support physical visits. For example, of the approximate 2.6 million visits to GBIF from 1 January 2016 to 1 October 2017, four of the top ten countries with the greatest visitorship came from



**Fig. 1 | Digitized information for *Clupeobarbus hypsolepis* (Daget, 1959) occurrences.** The data are based on doi:10.15468/dl.bicm4q accessed via [www.gbif.org](http://www.gbif.org) on 23 October 2017. This report collects high-resolution images, locality information and copies of the field notes for the original collection, and links to museums holding this species in their collections. Credit: GBIF ([www.gbif.org](http://www.gbif.org)).

mega-diverse countries in the Global South (India, Brazil, Mexico and Colombia). Digitizing collections provides access to global biodiversity data regardless of means, physical location, or ability status, and visitorship to global portals indicates that those portals are reaching a global audience.

Lastly, digitization of collections can augment both formal and informal science education at the secondary and tertiary levels. Digital technology, in general, can increase diversity within the classroom<sup>5,6</sup>. By incorporating digitized collections into classrooms, teachers from around the world can now use these data to structure enquiry-driven classes, which help students learn the system in question while also gaining fundamental insights into the practice of biology. Informal science educators also use these collections to augment local education programmes<sup>7</sup>. While these opportunities were previously accessible to only a small subset of universities that hosted museums, these kinds of learning experiences are broadly available and Internet traffic on museum collection sites continues to rise (Fig. 1)<sup>7</sup>.

Despite such potential, there are still significant barriers to full engagement with digitized collections. First, reliable

Internet connections remain elusive in some areas, limiting people's access to digitized collections — especially those that include large file sizes (high-resolution images, geographic information system shapefiles and so on). Second, despite substantial funding from some governmental sources, digitization is expensive and time consuming. By some estimates, digitizing all of the world's natural history collections would take more than a millennium<sup>8</sup>. While automation of this process is an obvious improvement, it still necessitates costly human-mediated quality control. In fact, several citizen science projects are centred on this, such as Notes from Nature ([www.notesfromnature.org](http://www.notesfromnature.org)), a crowd-sourced portal where volunteers help to input digitized field labels into a machine-readable format. Additionally, many databases are still taxon- or region-specific, limiting the ability of researchers to look across multiple axes simultaneously. Lastly, there is a limiting lack of long-term financial support for these digitization efforts. Large-scale digitization projects such as the National Science Foundation's Advancing Digitization of Biodiversity Collections have provided financial support for individual projects since 2011, but to ensure that there

is a consistent and coherent international approach to digitization it is imperative to establish long-term, globally accessible funding programmes.

If we truly want to use digitization to help diversify our community, we need to move beyond merely making those digitized products accessible. True integration does not come simply through the presentation of material, but rather from meaningful collaborations across institutions and nations. A clear next step is prioritizing methods to meaningfully involve partners in a dialogue about best practice for implementing these collaborations. Providing a URL to a digitized image is not enough; true integration — and

diversity — comes from collaborative research and training programmes, where hypotheses, analyses and publication credit are shouldered by all members of the collaboration. 

**Joshua A. Drew<sup>1,2\*</sup>, Corrie S. Moreau<sup>3</sup> and Melanie L. J. Stiassny<sup>2</sup>**

<sup>1</sup>*Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, NY, USA.* <sup>2</sup>*Division of Vertebrate Zoology, Department of Ichthyology, American Museum of Natural History, New York, NY, USA.*

<sup>3</sup>*Department of Science and Education, Integrative Research Center, Field Museum of Natural History, Chicago, IL, USA.*

\*e-mail: [j.drew@columbia.edu](mailto:j.drew@columbia.edu)

Published online: 13 November 2017

<https://doi.org/10.1038/s41559-017-0401-6>

#### References

1. Campbell, L. G., Mehtani, S., Dozier, M. E. & Rinehart, J. *PLoS ONE* **8**, e79147 (2013).
2. Drew, J. A., Amatangelo, K. L. & Hufbauer, R. A. *PLoS ONE* **10**, e0140682 (2015).
3. Ferrer-Paris, J. R., Sánchez-Mercado, A., Rodríguez-Clark, K. M., Rodríguez, J. P. & Rodríguez, G. A. *Biol. Conserv.* **173**, 133–143 (2014).
4. Paknia, O., Sh, H. R. & Koch, A. *Org. Divers. Evol.* **15**, 619–629 (2015).
5. Drew, J. *Integr. Comp. Biol.* **55**, 926–932 (2015).
6. Monfils, A. K. et al. *Southeast. Nat.* **16**, 47–57 (2017).
7. Cook, J. et al. *BioScience* **64**, 725–734 (2014).
8. Blagoderov, V. et al. *ZooKeys* **209**, 133–146 (2012).

#### Competing interests

The authors declare no competing financial interests.