

Digitize types and introduce standard morphometry protocols in taxonomy

While describing a new species, it is often considered a good practice to examine and compare the types of all the congeners, which sometimes is made mandatory by reviewers. This often requires visitation to the repositories/natural history museums where the types of the congeners are deposited. Although taxonomically it is prudent to do so, practically it is not possible at times as the researcher may need to travel to many repositories which are sometimes spread across many countries. This is not easy due to the high travel cost, which might not be feasible for many researchers. Moreover, with the current COVID pandemic, such extensive travel is now, at best, avoidable. Therefore, can we think of a way to resolve this issue?

A prudent way would be to digitize all the type specimens deposited across all the repositories of the world. This would require high-resolution imageries, detailed morphometric measurements and detailed morphological re-description of the type specimens by the concerned repositories/natural history museums, particularly the old types which have been briefly described. However, all these measures must be undertaken following a uniform and standard protocol for photographing the types and standardized morphometric measurements.

The current digitization efforts of natural history collections are random, patchy and not coordinated on a global scale¹. Besides, random digitization of all-natural history collections would be a mammoth task, given that more than half a billion preserved specimens are deposited across such repositories/natural history museums². Of the 2 million-odd species described by science, there are about 5–6 million name-bearing types, including the synonyms³. So, any effort towards digitizing natural history collections should first involve the types. For that, the International Code of

Zoological Nomenclature and the International Code of Nomenclature for algae, fungi and plants need to develop standard protocols for each of the major taxonomic groups, after detailed deliberations, keeping in mind both the quantitative and qualitative data that need to be generated.

Another benefit of digitizing the types will be the potential use of these data in machine learning for the rapid identification of species. We are in the age of big data². Gemeinholzer *et al.*⁴ have deliberated on the potential of this approach; however, at the same time, rueing the fact that the taxonomic data currently generated are nowhere near the level at which machine learning needs to make sense of them.

Once a coordinated digitization of types begins, the question of hosting the database will arise. Wheeler *et al.*³ have proposed an open-access global e-type archive; however, the maintenance of such a database would require a dedicated team of programmers, technicians, data centres and regular working capital and intermittent capital expenditure. Instead, access to the respective database can be kept with the concerned repositories/natural history museums which can be managed with minor additions of resources to the existing workforce with limited capital expenditure requirements. The concerned repositories/natural history museums can even monetize access to such a database with a nominal fee.

If such a coordinated approach is undertaken by all the repositories/natural history museums of the world, then the need for personal examination of types may not be essential in many cases. It is understandable that with this mostly the quantitative data will be available; while for qualitative data, a researcher might still need to examine the types personally. However, to generate the detailed qualitative data, micro-computed

tomography and confocal laser scanning microscopy can also be used to generate 3D imageries².

Taxonomy is a basic and excellent tool for biological sciences. To access the giant database generated in taxonomy, a large number of online resource pools are available like ZooBank (<http://zoobank.org>), Catalogue of Life (<https://www.catalogueoflife.org>), Encyclopedia of Life (<https://eol.org>), Global Biodiversity Information Facility (<https://www.gbif.org>), etc. However, such information databases are neither comprehensive nor detailed. Although these online resource pools are the right steps in earnest, the current impetus and approach must be to digitize the types and introduce standard morphometry protocols. This will help to harness the power of the large number of natural history collections and the use of big data to make taxonomy more efficient and quick. These would result in expediting the documentation of the estimated 10 million species⁵, of which majority are unknown to science and in all probability are subjected to extinction pressure.

1. Berendsohn, W. G. and Seltmann, P., *Biodivers. Inform.*, 2010, **7**, 120–129.
2. Short, A. E. Z., Dikow, T. and Moreau, C. S., *Annu. Rev. Entomol.*, 2018, **63**, 513–530.
3. Wheeler, Q. *et al.*, *ZooKeys*, 2012, **209**, 193–202.
4. Gemeinholzer, B. *et al.*, *Org. Divers. Evol.*, 2020, **20**, 1–8.
5. Sweetlove, L., *Nature*, 2011; <https://doi.org/10.1038/news.2011.498>.

BHASKAR SAIKIA
BIKRAMJIT SINHA*

*Zoological Survey of India,
North Eastern Regional Centre,
Shillong 793 003, India
e-mail: bj.sinha@gov.in