Modeling Taxon Concepts: A new approach to an old problem

Richard L. Pyle[‡], Nicolas Bailly[§], David Remsen[|]

‡ Bishop Museum, Honolulu, Honolulu, United States of America
§ University of British Columbia / Beaty Biodiversity Museum, Vancouver, Canada
| Marine Biological Laboratory, Woods Hole, United States of America

Corresponding author: Richard L. Pyle (deepreef@bishopmuseum.org)

Abstract

Although the biodiversity informatics community has recognized and understood the complexity of modeling information about scientific names and associated taxonomic concepts for more than three decades, many of the original questions and problems remain unresolved today. Because most biodiversity data is anchored to scientific names, and these names are governed by Codes of nomenclature, most effort and progress has focused on data structures centered around scientific names, rather than taxonomic concepts. But, as has been well documented in biodiversity data standards communities (e.g., Berendsohn (1995), Patterson et al. (2010), Pyle et al. (2021)), the relationship between the text-string scientific-name labels and the circumscribed conceptual taxa they are intended to represent is highly imprecise. Many attempts have been made to develop data models to represent taxonomic concepts as discrete, identifiable units to which biodiversity data can be linked. However, none has gained wide-spread adoption, often due to inherent subjective interpretations and the degree of taxonomic expertise required to define and interpret the individual units – aspects that limit their practical scalability. Similarly, previous efforts to develop taxon concept data models conflate properties of circumscription, classification, and nomenclature, resulting in overloaded notions of taxa that quickly become intractable. We describe an approach that mirrors centuries of actual taxonomic practice, rooted in fundamental properties of Code-regulated scientific names, which can leverage sources of existing digital information to represent taxonomic concepts in a highly structured, objective and computable way. It isolates the properties of circumscription from those of classification and nomenclature, but enables algorithmic integration of these three separate facets of taxonomic information using consistent informatic structures.

Keywords

Presenting author

Richard L. Pyle

Presented at

TDWG 2022

Conflicts of interest

References

- Berendsohn W (1995) The concept of "potential taxa" in databases. TAXON 44 (2): 207-212. https://doi.org/10.2307/1222443
- Patterson DJ, Cooper J, Kirk PM, Pyle RL, Remsen DP (2010) Names are key to the big new biology. Trends in Ecology & Evolution 25 (12): 686-691. https://doi.org/10.1016/j.tree.2010.09.004
- Pyle R, Barik S, Christidis L, Conix S, Costello MJ, van Dijk PP, Garnett S, Hobern D, Kirk P, Lien A, Orrell T, Remsen D, Thomson S, Wambiji N, Zachos F, Zhang Z, Thiele K (2021) Towards a global list of accepted species V. The devil is in the detail. Organisms Diversity & Evolution 21 (4): 657-675. https://doi.org/10.1007/s13127-021-00504-0