Extended Taxonomic Curation: Moving beyond species lists to linking species data

Nathan S. Upham‡, Caleb Powell‡, Laura Rocha Prado‡, Nico Franz‡, Beckett Sterner‡

‡ Arizona State University, Tempe, United States of America

Corresponding author: Nathan S. Upham (nathan.upham@asu.edu), Beckett Sterner (beckett.sterner@asu.edu)

Abstract

Taxonomy is at the center of modern biodiversity science. No species can be systematically studied until it is defined, and no observation can be linked to related data without a taxonomic label. However, taxonomy is also a science in constant flux—even well-studied groups like Mammalia have fluctuated by >25% in recognized species in the last decade (Burgin et al. 2018, MDD 2022a, MDD 2022b). As a result, there are calls to create a “global list of accepted species” to increase taxonomic stability, particularly for policy decisions in biodiversity conservation and management (Garnett et al. 2020). The counterargument notes that forcing definitional consensus is likely to further inequities, and that a pluralistic, coordinated approach to taxonomy can be achieved with innovative cyberinfrastructure designs and services (Sterner et al. 2020, Franz and Sterner 2018).

Here, we propose that digitally “extended” taxonomic curation can play new and innovative roles in

1. linking observational data to alternative taxonomic concepts; and
2. enabling fit-for-use taxonomy to inform policy decisions.

Taxonomic curators (TCs) have traditionally limited their activities to making lists of accepted species and higher taxa. However, most of today's biodiversity questions require observational data (e.g., specimen occurrences) that are taxonomically coherent, not just name lists, and for those linked data to be digitally available in public databases. If the collective activities of TCs can be effectively unified across distributed networks, they might facilitate the transition to Extended Specimen Networks of taxonomically coherent biodiversity data, a core goal of current research initiatives (e.g., Lendemer et al. (2020)).

Beyond lists of species names is the domain of what names mean in practice (i.e., taxonomic concepts), which often differs by author (Fig. 1). Here we argue that curating the various lines of evidence that represent taxonomic concepts—what we call Species Meaning Artifacts (SMArts)—is a promising strategy for keeping track of how species splits and lumps will affect observational data records in the Global Biodiversity Information Facility (GBIF) or National Center for Biotechnology Information (NCBI). Instead of labeling

records by a static name, records can be digitally associated with SMArt evidence from alternative taxonomies (e.g., geographic range maps before/after a species split). Networks of TCs curating digital SMArts will enable 'taxonomically intelligent' data aggregation (Bisby 2000), a long-pursued goal in biodiversity data science that, once realized, promises to enable investigations ranging from viral spillover to biodiversity loss (Upham et al. 2021).

Keywords

biodiversity data science, mammals, taxonomic intelligence

Presenting author

Nathan S. Upham

Presented at

TDWG 2022

Funding program

NIH NIAID grant 1R21AI164268-01 ("Intelligently predicting viral spillover risks from bats and other wild mammals")

Conflicts of interest

References

- Bisby F (2000) The Quiet Revolution: Biodiversity Informatics and the Internet. Science 289 (5488): 2309-2312. https://doi.org/10.1126/science.289.5488.2309
- Burgin C, Colella J, Kahn P, Upham N (2018) How many species of mammals are there? Journal of Mammalogy 99 (1): 1-14. https://doi.org/10.1093/jmammal/gyx147
- Franz N, Sterner B (2018) To increase trust, change the social design behind aggregated biodiversity data. Database 2018 https://doi.org/10.1093/database/bax100
- Garnett S, Christidis L, Conix S, Costello M, Zachos F, Bánki O, Bao Y, Barik S, Buckeridge J, Hobern D, Lien A, Montgomery N, Nikolaeva S, Pyle R, Thomson S, Dijk PPv, Whalen A, Zhang Z, Thiele K (2020) Principles for creating a single authoritative list of the world's species. PLOS Biology 18 (7). https://doi.org/10.1371/journal.pbio.3000736
- Lendemer J, Thiers B, Monfils AK, Zaspel J, Ellwood ER, Bentley A, LeVan K, Bates J, Jennings D, Contreras D, Lagomarsino L, Mabee P, Ford LS, Guralnick R, Gropp RE,

Revelez M, Cobb N, Seltmann K, Aime MC (2020) The Extended Specimen Network: A Strategy to Enhance US Biodiversity Collections, Promote Research and Education. BioScience 70 (1): 23-30. https://doi.org/10.1093/biosci/biz140

- MDD (2022a) Mammal Diversity Database (zenodo). DOI: 10.5281/zenodo.4139722 Type: dataset. https://doi.org/10.5281/zenodo.4139722
- MDD (2022b) Mammal Diversity Database (webpage). This is a real-time upload of the MDD v.1.9 taxonomy published 1 April 2022 on the mammaldiversity.org website.. URL: https://www.mammaldiversity.org/
- Sterner B, Witteveen J, Franz N (2020) Coordinating dissent as an alternative to consensus classification: insights from systematics for bio-ontologies. History and Philosophy of the Life Sciences 42 (1). https://doi.org/10.1007/s40656-020-0300-z
- Upham N, Poelen J, Paul D, Groom Q, Simmons N, Vanhove MM, Bertolino S, Reeder D, Bastos-Silveira C, Sen A, Sterner B, Franz N, Guidoti M, Penev L, Agosti D (2021) Liberating host–virus knowledge from biological dark data. The Lancet Planetary Health https://doi.org/10.1016/S2542-5196(21)00196-0

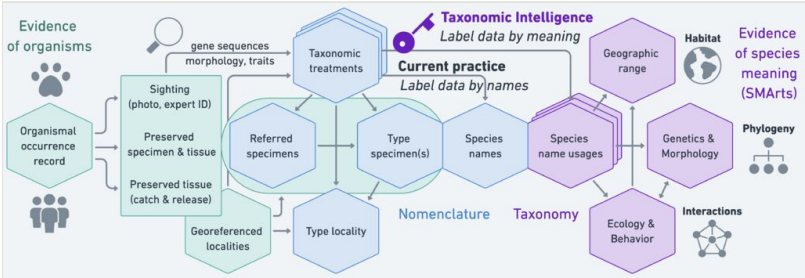

Figure 1.

Knowledge of species is rooted in organismal observations, and created via the flow of information from nomenclature (naming species) to taxonomy (defining species boundaries and relationships). Current practice is to label observational data by names alone; however, curating the lines of evidence that represent the conceptual meanings of those names, and how those meanings differ among authors through time, will allow for more accurate data labeling and aggregation (i.e., taxonomic intelligence).