

Elevating the Fitness of Use of GBIF Occurrence Datasets: A proposal for peer review

Vijay Barve ‡, §

‡ Natural History Museum of Los Angeles County, Los Angeles, California, United States of America

§ Nature Mates Nature Club, Kolkata, India

Corresponding author: Vijay Barve (vijay.barve@gmail.com)

Abstract

Biodiversity data plays a pivotal role in understanding and conserving our natural world. As the largest occurrence data aggregator, the Global Biodiversity Information Facility ([GBIF](#)) serves as a valuable platform for researchers and practitioners to access and analyze biodiversity information from across the globe (Ball-Damerow et al. 2019). However, ensuring the quality of GBIF datasets remains a critical challenge (Chapman 2005).

The community emphasizes the importance of data quality and its direct impact on the fitness of use for biodiversity research and conservation efforts (Chapman et al. 2020). While GBIF continues to grow in terms of the quantity of data it provides, the quality of these datasets varies significantly (Zizka et al. 2020). The biodiversity informatics community has been working diligently to ensure data quality at every step of data creation, curation, publication (Waller et al. 2021), and end-use (Gueta et al. 2019) by employing automated tools and flagging systems to identify and address issues. However, there is still more work to be done to effectively address data quality problems and enhance the fitness of use for GBIF-mediated data.

I highlight a missing component in GBIF's data publication process: the absence of formal peer reviews. Despite GBIF encompassing the essential elements of a data paper, including detailed metadata, data accessibility, and robust data citation mechanisms, the lack of peer review hinders the credibility and reliability of the datasets mobilized through GBIF.

To bridge this gap, I propose the implementation of a comprehensive peer review system within GBIF. Peer reviews would involve subjecting GBIF datasets to rigorous evaluation by domain experts and data scientists, ensuring the accuracy, completeness, and consistency of the data. This process would enhance the trustworthiness and usability of datasets, enabling researchers and policymakers to make informed decisions based on reliable biodiversity information.

Furthermore, the establishment of a peer review system within GBIF would foster collaboration and knowledge exchange among the biodiversity community, as experts provide constructive feedback to dataset authors. This iterative process would not only improve data quality but also encourage data contributors to adhere to best practices, thereby elevating the overall standards of biodiversity data mobilization through GBIF.

Keywords

biodiversity data, data quality, Global Biodiversity Information Facility data publication

Presenting author

Vijay Barve

Presented at

TDWG 2023

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Ball-Damerow J, Brenskelle L, Barve N, Soltis P, Sierwald P, Bieler R, LaFrance R, Ariño A, Guralnick R (2019) Research applications of primary biodiversity databases in the digital age. PLOS ONE 14 (9). <https://doi.org/10.1371/journal.pone.0215794>
- Chapman A, Belbin L, Zermoglio P, Wieczorek J, Morris P, Nicholls M, Rees ER, Veiga A, Thompson A, Saraiva A, James S, Gendreau C, Benson A, Schigel D (2020) Developing Standards for Improved Data Quality and for Selecting Fit for Use Biodiversity Data. Biodiversity Information Science and Standards 4 <https://doi.org/10.3897/biss.4.50889>
- Chapman AD (2005) Principles of Data Quality. Global Biodiversity Information Facility. <https://doi.org/10.15468/doc.jrgg-a190>
- Gueta T, Barve V, Nagarajah T, Gibas P, Carmel Y (2019) Introducing 'The bdverse': a family of R packages for biodiversity data. Biodiversity Information Science and Standards 3 <https://doi.org/10.3897/biss.3.37643>
- Waller J, Volik N, Mendez F, Hahn A (2021) GBIF Data Processing and Validation. Biodiversity Information Science and Standards 5 <https://doi.org/10.3897/biss.5.75686>
- Zizka A, Antunes Carvalho F, Calvente A, Rocio Baez-Lizarazo M, Cabral A, Coelho JFR, Colli-Silva M, Fantinati MR, Fernandes M, Ferreira-Araújo T, Gondim Lambert Moreira F, Santos NMC, Santos TAB, dos Santos-Costa RC, Serrano F, Alves da Silva AP, de Souza Soares A, Cavalcante de Souza PG, Calisto Tomaz E, Vale VF, Vieira TL,

Antonelli A (2020) No one-size-fits-all solution to clean GBIF. PeerJ 8 <https://doi.org/10.7717/peerj.9916>