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)
Received: 05 Sep 2023 | Published: 07 Sep 2023
Citation: Barve V (2023) Elevating the Fitness of Use of GBIF Occurrence Datasets: A proposal for peer review. Biodiversity Information Science and Standards 7: e112237. https://doi.org/10.3897/biss.7.112237

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.

© Barve V. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
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**

• 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