BiodivBank: designing a global repository and portal for structured biodiversity data

Anahita J.N. Kazem[‡,§], Birgitta Koenig-Ries[§], Henrique M. Pereira[‡,|], Jonathan M. Chase[‡,|]

‡ German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany
§ Friedrich Schiller University, Jena, Germany
| Martin Luther University, Halle-Wittenberg, Germany

Corresponding author: Anahita J.N. Kazem (anahita.kazem@idiv.de)

Abstract

Our goal is to mobilize global species abundance and assemblage information, via a dedicated and openly accessible data repository and web portal. Preservation of raw observational data is a complement to the modelled geographic projections that are the focus of related projects such as the EBV Data Portal, which provides access to EBV (Essential Biodiversity Variables) raster datasets. Key to this effort is an emphasis on the use of standardized terms to ensure interoperability with other services, as well as annotation of (meta)data fields and values to aid discoverability and create a product that adheres to the FAIR data principles, requesting data to be Findable, Accessible, Interoperable, and Reusable (Wilkinson et al. 2016). For example, retrieval and storage of globally unique, persistent and resolvable identifiers for core aspects of the data such as taxon names and geospatial locations can serve not only to underpin more advanced search functionality for portal users, but also facilitate future efforts to link open data within online knowledge networks. However addressing social challenges may be equally important in determining the success of the portal. Where do we fit within the biodiversity data landscape and how can we incentivize researchers to submit datasets? What is our data policy? How should we retain provenance and ensure visibility for data providers? Sometimes the choice of a technical solution can itself lower social barriers, for example by providing an intuitive graphical interface to ease upload of the detailed methodological metadata characteristic of these event-based datasets, or by blurring geographic coordinates in order to reduce the risk that threatened species are located. We will present our plans for the web platform currently under construction, and welcome comment on them, to ensure that the portal meets the needs of a diverse range of potential stakeholders.

Keywords

species abundance, biodiversity informatics, open science, FAIR principles

Presenting author

Anahita J.N. Kazem

Presented at

TDWG 2022

Conflicts of interest

References

- Wilkinson M, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J, da Silva Santos LB, Bourne P, Bouwman J, Brookes A, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo C, Finkers R, Gonzalez-Beltran A, Gray AG, Groth P, Goble C, Grethe J, Heringa J, 't Hoen PC, Hooft R, Kuhn T, Kok R, Kok J, Lusher S, Martone M, Mons A, Packer A, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S, Schultes E, Sengstag T, Slater T, Strawn G, Swertz M, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3 (1). https://doi.org/10.1038/sdata.2016.18