ARPHA Conference Abstracts 8: e151692
doi: 10.3897/aca.8.e151692

OPEN ACCESS

Conference Abstract

Enhancing data integration with existing FAIR enabling resources: a case for using the Darwin Core and its extensions to integrate eLTER biosphere data with other biodiversity data

Kate Ingenloff ‡

‡ GBIF Secretariat, Copenhagen, Denmark

Corresponding author: Kate Ingenloff (kingenloff@gbif.org)

Received: 28 Feb 2025 | Published: 28 May 2025

Citation: Ingenloff K (2025) Enhancing data integration with existing FAIR enabling resources: a case for using the Darwin Core and its extensions to integrate eLTER biosphere data with other biodiversity data. ARPHA Conference Abstracts 8: e151692. https://doi.org/10.3897/aca.8.e151692

Abstract

Regional and global efforts to monitor the status of, and identify mitigating actions for, climate change impacts and biodiversity loss requires copious amounts of survey and monitoring data that can only be obtained through the integration and analysis of multiple independently derived primary datasets from across a broad range of disciplines. Despite mandates to make ecological data FAIR and open (Wilkinson et al. 2016), data scientists and researchers attempting to integrate data from heterogeneous sources continue to face a series of pre-analysis hurdles including: finding data fit-for-use, non-interoperability across heterogenous data types and sources, a lack of full openness of data and broad availability through established research institutes, and missing or incomplete machine readable metadata (Blair et al. 2019). While overcoming these hurdles is no small feat, mechanisms do exist to facilitate the process; and two, in particular, are readily implementable.

1. *Darwin Core biodiversity data standard.*--Adopting a community-accepted data standard and vocabulary such as the Darwin Core (DwC) biodiversity data standard (https://dwc.tdwg.org/; Wieczorek et al. 2012) addresses three big data hurdles at once. Data standards are the primary mechanism for putting the ‘I’

© Ingenloff K. 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.

(interoperable) in 'FAIR' data practices, ensuring machine readability of dataset metadata, and enhancing data discovery. Darwin Core is a community-maintained biodiversity data standard maintained by Biodiversity Information Standards (TDWG, https://www.tdwg.org/). The standard is constantly undergoing review and expansion to better accommodate and report facets of biodiversity data in a structured manner.

2. *Open data aggregators.*--Enhancing opportunities to increase findability of and access to monitoring data is another critical set of hurdles to overcome. While some large research groups and networks maintain their own data catalogues and repositories (e.g. eLTER https://b2share.eudat.eu/, NEON https://www.neonscience.org/data), other may rely on community-specific (e.g., EDI https://portal.edirepository.org/nis/home.jsp, PANGAEA https://www.pangaea.de/) or more general data repositories (e.g., Dryad, https://datadryad.org/stash) to share data. Navigating the many existing domain-specific and generic repositories, however, can be daunting, and data discovery isn't always easy nor are suitable data guaranteed to be fully accessible when found. Sharing Darwin Core standardized data through a fully open global data aggregator such as GBIF (Global Biodiversity Information Facility, https://www.gbif.org/) provides a means to improve data discovery and re-use (with citation!) beyond what is capable through traditional repositories alone.

For example, consider the European Long-Term Ecological Research (eLTER, https://elter-ri.eu/) network. To address the basic tenet that ecological research is a cross-disciplinary field, requiring data from multiple domains to address questions of broad relevance, eLTER implemented their Whole Systems Approach to In-Situ Research on Life-supporting Systems (WAILS; Mirtl et al. 2021). The result is a standardized, systematic collection of monitoring data representing five distinct research areas from across Europe. These data have the potential to play a critical role in facilitating research and big data modeling efforts (for example, digital twins) across a broad range of data communities.

Using an eLTER biosphere dataset integrated into the BioDT Grassland Biodiversity Dynamics prototype digital twin (https://biodt.eu/use-cases/grassland-biodiversity-dynamics) as a case study, I will introduce the Darwin Core biodiversity data standard as an effective and implementable mechanism for improving the FAIRness of biodiversity survey data and demonstrate how sharing DwC formatted biodiversity data through a large global data aggregator such as GBIF is a complementary approach to other data sharing practices that can enhance the visibility and re-use of existing survey and monitoring data.

Keywords

Data standards, data interoperability, data aggregator, data access, FAIR data, open data

Presenting author

Kate Ingenloff

Presented at

ORAL

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Blair G, Henrys P, Leeson A, Watkins J, Eastoe E, Jarvis S, Young P (2019) Data Science of the Natural Environment: A Research Roadmap. Frontiers in Environmental Science 7 https://doi.org/10.3389/fenvs.2019.00121
- Mirtl M, Kuhn I, Montheith D, Bäck J, Orenstein D, Provenzale A, Zacharias S, Haase P, Shachak M (2021) Whole System Approach for in-situ research on Life Supporting Systems in the Anthropocene (WAILS). EGU General Assembly 2021 https://doi.org/10.5194/egusphere-egu21-16425
- Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7 (1). https://doi.org/10.1371/journal.pone.0029715
- 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