

Conference Abstract

Best Practice for Publishing Environmental DNA (eDNA) Data According to FAIR Principles

Miwa Takahashi^{‡,§}, Tobias Guldberg Frøslev[‡], Gareth Jenkins[¶], Martin Laporte[#], Bettina Thalinger[¶], Christopher Jerde[«], Sachit Rajbhandari[»], Thomas Stjernegaard Jeppesen[‡], Andrew Bissett[»], Bruce Deagle[»], Erin Hahn[^], Joana Pauperio[^], Peggy Newman[‡], Katy Klymus[?], Oliver Berry[‡]

‡ Commonwealth Scientific and Industrial Research Organisation, Crawley, Australia

§ Curtin University, Bentley, Australia

| Global Biodiversity Information Facility, Copenhagen, Denmark

¶ John Wiley & Sons Ltd, Oxford, United Kingdom

Ministère de l'Environnement, de la Lutte contre les changements climatiques, de la Faune et des Parcs, Québec, Canada

» University of Innsbruck, Innsbruck, Austria

« Marine Science Institute, University of California, Santa Barbara, United States of America

» CSIRO, Hobart, Australia

^ CSIRO, Canberra, Australia

^ European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom

‡ Atlas of Living Australia, Canberra, Australia

? USGS, Columbia Environmental Research Center, Columbia, MO, United States of America

Corresponding author: Miwa Takahashi (miwa.takahashi@csiro.au)

Received: 24 Sep 2024 | Published: 14 Oct 2024

Citation: Takahashi M, Frøslev T, Jenkins G, Laporte M, Thalinger B, Jerde C, Rajbhandari S, Jeppesen T, Bissett A, Deagle B, Hahn E, Pauperio J, Newman P, Klymus K, Berry O (2024) Best Practice for Publishing Environmental DNA (eDNA) Data According to FAIR Principles. Biodiversity Information Science and Standards 8: e137742. <https://doi.org/10.3897/biss.8.137742>

Abstract

Reversing global biodiversity loss will require transformational human actions and robust measurements of their effectiveness. Diversity assessment using environmental DNA (eDNA) has emerged as a cutting-edge technique with the potential to address the challenges of measuring biodiversity. Vast amounts of eDNA sequences and eDNA-based species detections are generated in scientific studies. These datasets are typically stored in a variety of different repositories in multiple formats, hindering their reuse (Berry et al. 2020). Ensuring the publication of eDNA data following the FAIR (Findable, Accessible, Interoperable, Reusable) principles (Wilkinson et al. 2016) would

revolutionise environmental assessment, including monitoring of biodiversity, individual species, and interactions across extensive spatial and temporal scales, and generate critical knowledge for evidence-based management. Archiving FAIR eDNA data requires standardising data formats and vocabularies, cyberinfrastructures, guidelines, data sharing policy, and collaboration among scientists and institutions. Some of these requirements are addressed by existing data standards and infrastructures, including Darwin Core (DwC) (Wieczorek et al. 2012), Minimum Information about any (x) Sequence (MlxS) (Yilmaz et al. 2011), the Global Biodiversity Information Facility (GBIF) network, and International Nucleotide Sequence Database Collaboration (INSDC) partners (Arita et al. 2020).

However, multiple challenges remain and FAIR data practices have yet to be established among the eDNA community. This is partly because critical attributes unique to eDNA data are not adequately accommodated by existing standards. For example, monitoring contamination and excluding non-target taxa, and the parameters used for quality filtering and species detection vary greatly between studies, depending on the study scopes and the associated financial and ecological costs of incorrectly inferring presence or absence. Making such information FAIR is needed for future studies reusing data and requiring high confidence levels in species detection and taxonomic assignment. Furthermore, the procedures of targeted-taxon detection approach (e.g., interpretations of quantitative polymerase chain reaction (qPCR) results in detecting the presence of DNA from individual taxa) have not yet been fully captured by existing standards. Increasing efforts have been made to establish minimum reporting requirements to validate eDNA study methods and data (Klymus et al. 2020, Thalinger et al. 2021). These requirements need further development to be translated into data standards and formats to enhance machine readability and reusability, and to support and guide the eDNA community, so that they are effectively utilised.

In this talk, we share our best practice guide for formatting and publishing eDNA data, developed by an international multidisciplinary working-group comprising eDNA researchers, journal editors, and biodiversity and omics data scientists. We identified required data types, formats and metadata checklists through reviewing and integrating existing data standards, devising subject-specific vocabularies, and introducing additional terms to accommodate the distinctive properties of eDNA data. Implementing the FAIR eDNA data best practice guide, offers a pivotal step towards standardising and enhancing the publication and re-use of eDNA data.

Keywords

data management, data sharing, reproducibility, standardisation, metabarcoding, species-specific assay, species occurrence

Presenting author

Miwa Takahashi

Presented at

SPNHC-TDWG 2024

Funding program

Environomics Future Science Platform

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Arita M, Karsch-Mizrachi I, Cochrane G, et al. (2020) The international nucleotide sequence database collaboration. *Nucleic Acids Research* 49 <https://doi.org/10.1093/nar/gkaa967>
- Berry O, Jarman S, Bissett A, et al. (2020) Making environmental DNA (eDNA) biodiversity records globally accessible. *Environmental DNA* 3 (4): 699-705. <https://doi.org/10.1002/edn3.173>
- Klymus KE, Merkes CM, Allison MJ, et al. (2020) Reporting the limits of detection and quantification for environmental DNA assays. *Environmental DNA* 2 (3): 271-282. <https://doi.org/10.1002/edn3.29>
- Thalinger B, Deiner K, Harper LR, et al. (2021) A validation scale to determine the readiness of environmental DNA assays for routine species monitoring. *Environmental DNA* 3 (4): 823-836. <https://doi.org/10.1002/edn3.189>
- Wieczorek J, Bloom D, Guralnick R, et al. (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 JJ, et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3 (1). <https://doi.org/10.1038/sdata.2016.18>
- Yilmaz P, Kottmann R, Field D, et al. (2011) Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *Nature Biotechnology* 29 (5): 415-420. <https://doi.org/10.1038/nbt.1823>