


































## PERSPECTIVE OPEN ACCESS

# A Metadata Checklist and Data Formatting Guidelines to Make eDNA FAIR (Findable, Accessible, Interoperable, and Reusable)

Miwa Takahashi<sup>1,2</sup>  | Tobias Guldberg Frøslev<sup>3</sup>  | Joana Paupério<sup>4</sup>  | Bettina Thalinger<sup>5</sup>  | Katy Klymus<sup>6</sup>  | Caren C. Helbing<sup>7</sup>  | Cecilia Villacorta-Rath<sup>8</sup>  | Katherine Silliman<sup>9</sup>  | Luke R. Thompson<sup>9,10</sup>  | Sean P. Jungbluth<sup>9,10,11</sup>  | Suk Yee Yong<sup>12</sup>  | Stephen Formel<sup>13</sup>  | Gareth Jenkins<sup>14</sup>  | Martin Laporte<sup>15</sup>  | Bruce Deagle<sup>16</sup>  | Sachit Rajbhandari<sup>16</sup>  | Thomas Stjernegaard Jeppesen<sup>3</sup>  | Andrew Bissett<sup>17</sup>  | Christopher Jerde<sup>18</sup>  | Erin E. Hahn<sup>19</sup>  | Lynn M. Schriml<sup>20</sup>  | Christopher Hunter<sup>21</sup>  | Peggy Newman<sup>22</sup>  | Peter Woollard<sup>4</sup>  | Lynsey R. Harper<sup>23</sup>  | Nicholas Dunn<sup>23</sup>  | Katrina West<sup>1</sup>  | Rachel Haderlé<sup>24,25</sup>  | Shaun Wilkinson<sup>2,26</sup>  | Neha Acharya-Patel<sup>7</sup>  | Mark Louie D. Lopez<sup>7</sup>  | Guy Cochrane<sup>4</sup>  | Oliver Berry<sup>1</sup> 

<sup>1</sup>Commonwealth Scientific and Industrial Research Organisation, National Collections and Marine Infrastructure, Crawley, Western Australia, Australia | <sup>2</sup>Faculty of Science and Engineering, Curtin University, Bentley, Western Australia, Australia | <sup>3</sup>Global Biodiversity Information Facility, Universitetsparken, Copenhagen Ø, Denmark | <sup>4</sup>European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, UK | <sup>5</sup>Department of Zoology, University of Innsbruck, Innsbruck, Austria | <sup>6</sup>U.S. Geological Survey, Columbia Environmental Research Center, Columbia, Missouri, USA | <sup>7</sup>Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia, Canada | <sup>8</sup>Centre for Tropical Water and Aquatic Ecosystem Research (TropWATER), James Cook University, Townsville, Queensland, Australia | <sup>9</sup>Atlantic Oceanographic and Meteorological Laboratory, National Oceanic and Atmospheric Administration, Ocean Chemistry and Ecosystems Division, Miami, Florida, USA | <sup>10</sup>Northern Gulf Institute, Mississippi State University, Starkville, Mississippi, USA | <sup>11</sup>Estuary and Ocean Science Center, San Francisco State University, Tiburon, California, USA | <sup>12</sup>Commonwealth Scientific and Industrial Research Organisation, Information Management & Technology Scientific Computing, Eveleigh, New South Wales, Australia | <sup>13</sup>U.S. Geological Survey, Science Analytics and Synthesis, Lakewood, Colorado, USA | <sup>14</sup>John Wiley & Sons Ltd, Oxford, UK | <sup>15</sup>Institut de Biologie Intégrative et Des Systèmes (IBIS), Université Laval, Québec, Québec, Canada | <sup>16</sup>Commonwealth Scientific and Industrial Research Organisation, National Collections and Marine Infrastructure, Hobart, Tasmania, Australia | <sup>17</sup>Commonwealth Scientific and Industrial Research Organisation, Environment, Hobart, Tasmania, Australia | <sup>18</sup>Marine Science Institute, University of California, Santa Barbara, California, USA | <sup>19</sup>Commonwealth Scientific and Industrial Research Organisation, National Research Collections Australia, Canberra, Australian Capital Territory, Australia | <sup>20</sup>School of Medicine, University of Maryland, Baltimore, Maryland, USA | <sup>21</sup>GigaScience Press, Pak Shek Kok, Hong Kong, Hong Kong SAR | <sup>22</sup>Commonwealth Scientific and Industrial Research Organisation, National Collections and Marine Infrastructure, Melbourne, Victoria, Australia | <sup>23</sup>Natural England, Evidence Directorate, York, UK | <sup>24</sup>Muséum National D'histoire Naturelle, Institut de Systématique, Evolution, Biodiversité (ISYEB), Paris, France | <sup>25</sup>Station Marine de Dinard du Muséum National D'histoire Naturelle, Dinard, France | <sup>26</sup>Wilderlab, Wellington, New Zealand

**Correspondence:** Miwa Takahashi ([miwa.takahashi@csiro.au](mailto:miwa.takahashi@csiro.au); [miwatakahashi582@gmail.com](mailto:miwatakahashi582@gmail.com))

**Received:** 24 January 2025 | **Revised:** 20 March 2025 | **Accepted:** 21 March 2025

**Funding:** This work was supported by European Union HORIZON EUROPE Research and innovation programme (101057437); National Oceanic and Atmospheric Administration (NA21OAR4320190) and Commonwealth Scientific and Industrial Research Organisation.

**Keywords:** data management | data reuse | data sharing | metabarcoding | reproducibility | species occurrence | standardization | targeted assay

## ABSTRACT

The success of environmental DNA (eDNA) approaches for species detection has revolutionized biodiversity monitoring and distribution mapping. Targeted eDNA amplification approaches, such as quantitative PCR, have improved our understanding of species distribution, and metabarcoding-based approaches have enabled biodiversity assessment at unprecedented scales and taxonomic resolution. eDNA datasets, however, are often scattered across repositories with inconsistent formats, varying access restrictions, and inadequate metadata; this limits their interoperation, reuse, and overall impact. Adopting FAIR (Findable, Accessible, Interoperable,

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2025 The Author(s). *Environmental DNA* published by John Wiley & Sons Ltd.

and Reusable) data practices with eDNA data can transform the monitoring of biodiversity and individual species and support data-driven biodiversity management across broad scales. FAIR practices remain underdeveloped in the eDNA community, partly due to gaps in adapting existing vocabularies, such as Darwin Core (DwC) and Minimum Information about any (x) Sequence (MIxS), to eDNA-specific needs and workflows. To address these challenges, we propose a comprehensive FAIR eDNA (FAIRe) Metadata Checklist, which integrates existing data standards and introduces new terms tailored to eDNA workflows. Metadata are systematically linked to both raw data (e.g., metabarcoding sequences, Ct/Cq values of targeted qPCR assays) and derived biological observations (e.g., Amplicon Sequence Variant (ASV)/Operational Taxonomic Unit (OTU) tables, species presence/absence). Along with formatting guidelines, tools, templates, and example datasets, we introduce a standardized, ready-to-use approach for FAIR eDNA practices. Through broad collaboration, we seek to integrate these guidelines into established biodiversity and molecular data standards, promote journal data policies, and foster user-driven improvements and uptake of FAIR practices among eDNA data producers. In proposing this standardized approach and developing a long-term plan with key databases and data standard organizations, the goal is to enhance accessibility, maximize reuse, and elevate the scientific impact of these valuable biodiversity data resources.

## 1 | Need for FAIR eDNA Data

The FAIR principles were developed to improve the Findability, Accessibility, Interoperability, and Reusability of scholarly digital research objects such as data and scripts (Wilkinson et al. 2016) (refer to Table 1 for the list of abbreviations, applicable throughout). Alongside open data principles (Mandeville et al. 2021), these principles promote the public good and enhance the transparency and reproducibility of research. Furthermore, the FAIR principles increase the long-term value of data, generate further knowledge at speed, and serve as a cost-saving measure (Beck et al. 2020; Gomes et al. 2022; Hampton et al. 2013; Jenkins et al. 2023; Kim and Stanton 2016; Leigh et al. 2024; Sholler et al. 2019; Soranno and Schimel 2014; Whitlock 2011). Open and FAIR principles have been adopted worldwide by organizations including the Research Data Alliance (RDA, <https://www.rd-alliance.org>), the Committee on Data of the International Science Council (CODATA, <https://codata.org>) and the Future of Research Communications and e-Scholarship (FORCE11) (Jacobsen et al. 2020; Martone 2015; Wilkinson et al. 2016). FAIR practices, however, have not yet been fully established in the field of environmental DNA (eDNA), limiting the opportunities to unlock its full potential (Berry et al. 2021; Borisenko et al. 2024; Kimble et al. 2022; Shea et al. 2023).

Given increasing pressures on the environment and accelerated species loss (Ceballos et al. 2015; Diaz et al. 2019), FAIR biodiversity data can support robust, real-time biomonitoring, biodiversity inventories, and provide a better foundation for decision making. Reversing the global loss of biodiversity and the associated resources and services requires transformational human actions and robust measurements of the effectiveness of those actions (Berry et al. 2021; Diaz et al. 2019). Yet, environmental measurement and biodiversity assessment at large scale, with high speed and accuracy, is very challenging. Recently, biodiversity science has adopted eDNA to help address these challenges, noting the various advantages of eDNA including non-invasive sampling, broad taxonomic coverage, high sensitivity, accuracy, scalability, and efficiency (Goldberg et al. 2015; Ruppert et al. 2019; Thomsen and Willerslev 2015). eDNA is also effective in situations where conventional methods for biological observation are impractical or impossible, such as in turbid waters, inaccessible subterranean environments, and for detecting elusive or cryptic species (Hashemzadeh Segherloo et al. 2022; Matthias et al. 2021; Saccò et al. 2022; Villacorta-Rath et al. 2022; West et al. 2020; White

et al. 2020). As a result of these advantages, it has been widely implemented for biodiversity detection and monitoring, including as part of national and international strategies, for example, National Aquatic Environmental DNA Strategy (Goodwin et al. 2024; Kelly et al. 2024), Fisheries and Oceans Canada (Baillie et al. 2019), DNAqua-net (Blancher et al. 2022), and the Canadian Standards Association (Helbing and Hobbs 2019).

There are two primary approaches to detect organisms of interest in eDNA studies. The first approach, targeted assays (e.g., quantitative polymerase chain reaction (qPCR), digital PCR (dPCR)), uses taxon-specific primers and probes to detect the presence of DNA from specific taxa. This method typically targets a single taxon (e.g., species or genus) (but also refer to Tsuji et al. (2018) and Wilcox et al. (2020) for multiplex and high-throughput PCR) from which eDNA concentration may be inferred as a proxy for abundance or biomass of the focal taxon (Rourke et al. 2022). The second approach, metabarcoding assays, detects entire taxon assemblages from a sample by amplifying and sequencing genetic marker regions (so-called barcode regions) that are conserved across different taxa (Coissac et al. 2012; Hebert et al. 2003). This approach assigns large numbers of DNA sequences to diverse taxa, offering a broad overview of biodiversity within the sample. Both targeted and metabarcoding assay approaches can produce tremendous volumes of complex data. The exact methodologies employed within each approach can vary significantly (Coissac et al. 2012; Takahashi et al. 2023).

Metabarcoding sequences (amplicon sequence variants (ASVs) or operational taxonomic units (OTUs)) are usually taxonomically annotated by comparing them to DNA reference databases, such as those available in the Barcode of Life Data System (BOLD) (Ratnasingham and Hebert 2007), the International Nucleotide Sequence Database Collaboration (INSDC) (Arita et al. 2021), and the Genome Taxonomy Database (GTDB) (Parks et al. 2022). As reference databases are continuously improved, for example through genetic sequencing of vouchered specimens (Delrieu-Trottin et al. 2019; Stoeckle et al. 2020; Weigand et al. 2019), eDNA sequences may receive more precise taxonomic assignment upon reanalysis. The ability to improve in accuracy and precision with time is a feature of eDNA data that is rare among biodiversity data types. This potential is, however, dependent on machine-readable and high-quality metadata, underscoring the importance of compliance with the FAIR principles. Adhering

**TABLE 1** | Glossary of abbreviations.

Acronyms	Definitions
ABS	Access and benefit-sharing
ALA	Atlas of Living Australia
APIs	Application programming interfaces
ASV	Amplicon sequence variant
BOLD	Barcode of Life Data System
CARE	Collective benefit, Authority to control, Responsibility and Ethics
DDBJ	DNA Data Bank of Japan
dPCR	Digital Polymerase Chain Reaction
DSI	Digital Sequence Information
DwC	Darwin Core
eDNA	Environmental DNA
ENA	European Nucleotide Archive
FAIR	Findable, Accessible, Interoperable, Reusable
FAIRe	FAIR eDNA (in the context of the current initiative)
GBIF	Global Biodiversity Information Facility
GGBN	Global Genome Biodiversity Network
GSC	Genomic Standards Consortium
INSDC	International Nucleotide Sequence Database Collaboration
IPLCs	Indigenous Peoples and Local Communities
MDT	Metabarcoding Data Toolkit
MIDs	Multiplexing identifiers
MIEM	Minimum Information for eDNA and eRNA Metabarcoding
MIMARKS	Minimum Information about a Marker Sequence
MIMS	Minimum Information about Metagenome or Environmental Sequence
MIQE	Minimum Information for publication of Quantitative real-time PCR Experiments
MIxS	Minimum Information about any (x) Sequence
NCBI	National Center for Biotechnology Information
OBIS	Ocean Biodiversity Information System
OTU	Operational Taxonomic Unit
PCR	Polymerase Chain Reaction
qPCR	Quantitative Polymerase Chain Reaction
SRA	Sequence Read Archive

(Continues)

**TABLE 1** | (Continued)

Acronyms	Definitions
TDWG	Biodiversity Information Standards (formerly Taxonomic Databases Working Group)

to these will also expand the lifespan and scale of eDNA data, thereby maximizing its power for biodiversity monitoring and providing additional value to existing biodiversity data (Berry et al. 2021; Thompson and Thielen 2023). For instance, FAIR eDNA data can be reused to increase the temporal and spatial scale of species and community observations (Crandall et al. 2023), fill important spatial and taxonomic gaps in our knowledge of global biodiversity, facilitate early invasive species detection (e.g., Helbing and Hobbs 2019), and enhance endangered species monitoring. Furthermore, currently unassigned sequences can contribute meaningful ecological insights, such as the taxon-independent community index (Wilkinson et al. 2024) or in habitat classification and predictive modeling (e.g., Frøslev et al. 2023). Altogether, implementing FAIR principles for eDNA data can maximize its applications and conservation impact.

## 2 | Current State of eDNA Data Practices and Challenges to Achieving FAIR Standards

### 2.1 | Data Sharing Policies

Data sharing is widely regarded as vital for ensuring reproducibility and has been adopted by the majority of ecological researchers (Herold 2015). The current standard practice for eDNA datasets is that upon publication of an associated manuscript, they are stored in digital repositories where they can be accessed by other researchers (Abarenkov et al. 2023; Leigh et al. 2024; Shea et al. 2023), but there is little oversight into which data are archived and how accessible they are. The establishment of public data archiving policies and requirements by scientific journals has been a key contributor to advancing data sharing practices (Sholler et al. 2019; Thrall et al. 2023; Whitlock 2011; Whitlock et al. 2010). Open-access journals such as GigaScience and GigaByte further support these efforts by providing dedicated data repositories (GigaDB) to ensure that publications are accompanied by supporting data that uphold principles of transparency, reproducibility, and FAIR standards (Armit et al. 2022). Additionally, funding agencies have also effectively implemented data sharing policies at national and international levels in various countries (“Australian Research Council,” 2021; “Open Research Europe | Open Access Publishing Platform,” 2025; “Proposal & Award Policies & Procedures Guide (PAPPG) (NSF 24-1) | NSF – National Science Foundation,” 2024; Hutchison et al. 2024).

These proactive approaches have led to a broader adoption of open data practices and a substantial increase in the proportion of published articles sharing their data. For instance, when journals mandate data archiving, there is a nearly 1000-fold increase in the ability to find data online compared to data in journals without a policy (Vines et al. 2013). Despite this progress, challenges to full FAIR principles compliance remain as the process is not yet fully standardized or

unambiguously defined. For example, many datasets fall short due to inconsistent formatting and sparse metadata, limiting their interoperability and reusability (Hassenr ck et al. 2021; Roche et al. 2015; Tedersoo et al. 2021). This may be due to a lack of awareness and familiarity around data standardization, existing guidelines, and the benefits of FAIR data sharing. However, existing FAIR guidelines for standardized formatting and sharing can be dissonant. Moreover, current metadata checklists (i.e., lists of data terms required and recommended for describing methods and datasets) implemented in standards are not always suitable for documenting the entire eDNA workflow. Nevertheless, each dataset published without meeting the FAIR principles can present a lost opportunity to preserve rich biodiversity data for reuse in research and decision making.

## 2.2 | Data Standards

The establishment and adoption of standardized data formats, vocabularies, and comprehensive metadata checklists provide a robust foundation for ensuring that archived data are machine readable, interoperable, and suitable for reuse (Vines et al. 2013). Darwin Core (DwC, [dwc.tdwg.org](http://dwc.tdwg.org)), developed by Biodiversity Information Standards (TDWG, [tdwg.org](http://tdwg.org)) and Minimum Information about any (x) Sequences (MIxS, [genomicsstandardsconsortium.github.io/mixs](http://genomicsstandardsconsortium.github.io/mixs)), developed by the Genomic Standards Consortium (GSC, [gensc.org](http://gensc.org)), are data standards to facilitate FAIR practices for biodiversity and DNA sequence data (Wieczorek et al. 2012; Yilmaz et al. 2011). Widely adopted by major data infrastructures like the INSDC ([www.insdc.org](http://www.insdc.org)) (Arita et al. 2021) and the Global Biodiversity Information Facility (GBIF, [www.gbif.org](http://www.gbif.org)), these standards can be powerful enablers of FAIR data practices. Furthermore, interoperability between DwC and MIxS has been established by a mapping of MIxS terms to a DwC extension (Meyer et al. 2023). The DNA-derived data extension to DwC was developed to accommodate biodiversity data in DwC-based biodiversity platforms like GBIF. The extension is currently maintained by GBIF (<http://rs.gbif.org/terms/1.0/DNADerivedData>) and consists of terms from MIxS, the Global Genome Biodiversity Network (GGBN) standard ([https://www.ggbn.org/ggbn\\_portal](https://www.ggbn.org/ggbn_portal)), the Minimum Information for publication of Quantitative real-time PCR Experiments (MIQE) Guidelines (Bustin et al. 2009), and a few custom terms (e.g., atomised versions of the primer terms on MIxS) (Abarenkov et al. 2023).

Despite dynamic community-driven development (Abarenkov et al. 2023; Bustin et al. 2009; Meyer et al. 2023; Wieczorek et al. 2012; Yilmaz et al. 2011), these existing data standards lack unambiguous and comprehensive terms, vocabularies, and guidance to fully capture some of the distinctive features of eDNA data and workflows. For instance, qPCR-based eDNA workflows, first introduced around 2010 (e.g., Ficetola et al. 2008; Thomsen et al. 2012), have been increasingly used for targeted assay detection (Takahashi et al. 2023). The most comprehensive qPCR checklist (i.e., the MIQE Guidelines) was, however, developed prior to the prevalence of eDNA studies and may need to be revised and updated to align with FAIR principles for qPCR-based eDNA reporting. Furthermore, procedures

for monitoring contamination and excluding nontarget taxa, and the parameters used for quality filtering and species detection vary greatly between studies, depending on the project scope and the associated financial and ecological costs of false positives and false negatives (Ficetola et al. 2015; Takahashi et al. 2023). Making detailed workflow descriptions FAIR can enhance future studies' ability to reuse data and ensure high confidence levels in species detection and taxonomic assignment. Increasing efforts have been made to establish minimum reporting requirements to validate eDNA study methods and data (Abbott et al. 2021; Borisenko et al. 2024; Gagn e et al. 2021; Kimble et al. 2022; Klymus et al. 2020; Thalinger et al. 2021). As a next step, these requirements could be translated into data standards and formats to enhance human and machine readability, interoperability, and reusability.

Like other machine-derived data, eDNA datasets are comprised of a variety of data components representing degrees of processing and refinement. Therefore, defining the essential data components to be shared is crucial to ensure that shared data are standardized, complete, and reusable. Components of eDNA datasets can be broadly grouped into three categories: (1) raw data, including amplification data (i.e., Ct/Cq values from targeted assay detection studies) and raw DNA sequences (i.e., FASTQ files from metabarcoding studies), (2) interpreted data, including estimated DNA copy numbers from standard curves and detections (1/0) of targeted taxa, raw and curated ASV/OTU tables, and taxonomic assignment of each ASV/OTU, and (3) metadata describing data content, provenance, methods, and associated environmental data for each project and sample. Each data type serves a different purpose and possesses distinct qualities. For example, raw data are essential for documentation and reproducibility, as well as for enabling future re-analysis with different or updated bioinformatic pipelines. In contrast, interpreted data and metadata provide immediate value for broader biodiversity research and data-driven decision-making. Sharing all these data components in standardized formats is crucial to making data FAIR.

## 2.3 | Cyberinfrastructure

There are a range of digital repositories available for depositing eDNA datasets, eliminating the need for a new or single repository. Instead, fostering collaborations among databases could ensure that shared eDNA datasets remain interoperable and reusable. These repositories can broadly be categorized into sequence databases and biodiversity databases. The INSDC sequence databases, which are a collaboration between the US National Center for Biotechnology Information (NCBI, Sayers et al. 2024), the European Nucleotide Archive (ENA, [www.ebi.ac.uk/ena](http://www.ebi.ac.uk/ena)) (Yuan et al. 2024), and the DNA Data Bank of Japan (DDBJ, [www.ddbj.nig.ac.jp](http://www.ddbj.nig.ac.jp)) (Kodama et al. 2024), aim to compile and disseminate DNA sequence data and associated metadata generated worldwide (Arita et al. 2021). Data archived in NCBI, ENA, and DDBJ are routinely exchanged and synchronized among these databases. Biodiversity databases such as GBIF, on the other hand, serve as an open-access online platform for global biodiversity data, with a primary focus on species occurrences. The types of data available in GBIF include digitized specimens from natural history collections,

observations from citizen science platforms (e.g., eBird ([ebird.org](http://ebird.org)), iNaturalist ([inaturalist.org](http://inaturalist.org))), and DNA-associated records (e.g., barcoded specimens from the BOLD database (Ratnasingham and Hebert 2007) and records from INSDC (Arita et al. 2021)), as well as metabarcoding and metagenomic datasets from individual studies (e.g., Shea and Boehm 2024) and initiatives such as MGnify (Richardson et al. 2023) (refer to Abarenkov et al. 2023). Other biodiversity databases based on the Darwin Core standard for biodiversity data, such as the Ocean Biodiversity Information System (OBIS, [obis.org](http://obis.org)) and the Atlas of Living Australia (ALA, [www.ala.org.au](http://www.ala.org.au)) focus on distinct ecosystems or regions and routinely exchange datasets. Given that eDNA data encompass both biodiversity information (e.g., inferred species occurrences) and molecular sequence data (e.g., raw reads), an optimal data standard would allow for describing eDNA datasets intended to be shared through both biodiversity and nucleotide sequence databases.

Publishing eDNA data in established databases can be a simple and powerful step toward achieving FAIR eDNA, as these platforms provide persistent sample and sequence identifiers across datasets, and offer robust, indexed search functions based on user-defined parameters, such as geographical regions, ecosystem types, taxa and gene regions, and making data highly findable and accessible. Moreover, eDNA data shared in biodiversity databases like GBIF and OBIS contribute to the global biodiversity records by adding DNA-derived species occurrences alongside data from other sources. These databases do not accept all three primary eDNA data components (metadata, raw data and interpreted data) as a whole, and therefore, require linkage if data are to be accessible for re-use. For example, GBIF focuses on inferred biodiversity data, such as detected species or taxa with spatiotemporal metadata, and representative curated DNA sequences. It, however, does not host raw DNA sequencing files (i.e., FASTQ files), but may link to the corresponding INSDC records. Conversely, INSDC databases host raw sequence data, but neither the inferred biodiversity data in an interoperable format nor eDNA-derived species occurrence data without DNA sequences, such as qPCR-based detections. Collaboration efforts between GBIF and INSDC (through ENA) are in progress to ensure cross-linking of related data resources and increase interoperability between the two data infrastructures (“Cross-infrastructure collaboration with ENA improves processing, quality of DNA-derived occurrences” 2021).

In contrast, eDNA practitioners can deposit their eDNA data as a whole in open data repositories, such as Dryad (<https://datadryad.org>), Zenodo (<https://zenodo.org>), or FigShare (<https://figshare.com>), which accept unrestricted data types. Data in these repositories, however, are not fully FAIR, as they are not standardized by default and are not indexed to a degree where they are functionally findable and interoperable. Hence, data reusers are required to manually access and reprocess the data before it can be re-analyzed and/or combined with other datasets. As a consequence, eDNA data published in these repositories are at risk of getting lost and becoming “dark data” (Heidorn 2008).

More recently, various infrastructures have been developed specifically to present eDNA data from particular regions, institutions, or taxa, for example, All Nippon eDNA Monitoring Network (ANEMONE, [db.anemone.bio](http://db.anemone.bio)), Wilderlab ([\[rlab.co.nz/explore\]\(http://rlab.co.nz/explore\)\), eDNAAtlas \(\[www.fs.usda.gov/research/rmrs/projects/ednatlas\]\(http://www.fs.usda.gov/research/rmrs/projects/ednatlas\)\), \(Young et al. 2018\)\), eDNA Explorer \(\[www.ednaexplorer.org\]\(http://www.ednaexplorer.org\)\), the Australian Microbiome \(\[www.australianmicrobiome.com\]\(http://www.australianmicrobiome.com\)\), Swedish ASV portal \(<https://asv-portal.biodiversitydata.se>\), ASV Board \(<https://asv.bolgermany.de/metabarcoding>\), Nonindigenous Aquatic Species \(NAS\) database \(<https://nas.er.usgs.gov/eDNA>\), and the global microbiome database under the National Microbiome Database Collaborative \(NMDC, \[data.microbiomedata.org\]\(http://data.microbiomedata.org\)\). These databases utilize custom data structures to serve their particular focus and functionalities, sometimes in combination with terms from more widely established data standards like MIXS and DwC. Publishing data from such databases to global databases \(e.g., INSDC and GBIF\) requires mapping custom data terms to an open data standard, which can be a lossy process but serves the purposes of discoverability and inclusion in the default global data pools.](http://www.wilde</a></p>
</div>
<div data-bbox=)

While the aforementioned databases and infrastructures share the goal to promote open and data-driven science, they also cater to specific focuses and requirements. Consequently, a spectrum of repositories exists, ranging from those accepting any research data in unrestricted formats (e.g., Dryad, Zenodo, FigShare) to those exclusively processing eDNA-derived data from a single source using standardized protocols (e.g., ANEMONE and Wilderlab). Because of the numerous repositories, there is likely not a need for a new or single repository. Instead, fostering collaborations among eDNA communities, databases, data standard organizations, and journal editors and publishers could ensure that shared eDNA datasets are comprehensive, interoperable, and reusable. This involves integrating eDNA data into existing databases and standards and mapping it across past, current, and future standards. This approach retains flexibility and interoperability with existing repositories while supporting the formatting of shared datasets into common standards. This, in turn, facilitates data mobilization, collation, and reuse.

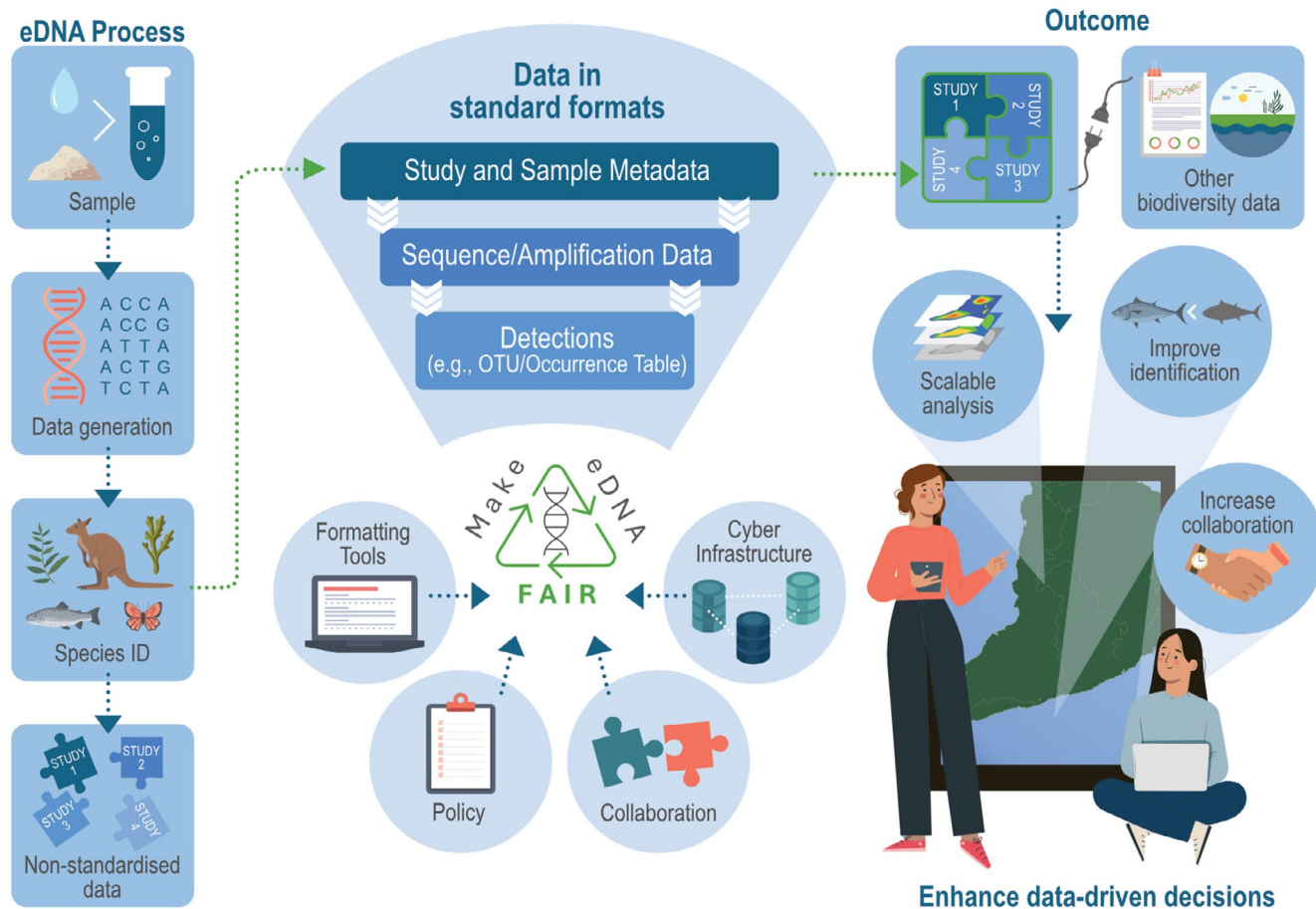
### 3 | Project Scope

Fine-tuning, optimizing, and extending existing data and metadata standards specifically for eDNA-based studies (both targeted and metabarcoding detection approaches) is an important first step toward achieving FAIR eDNA.

To this end, we have established an approach with the following steps:

1. Formation of an international, multidisciplinary working group comprising eDNA practitioners, journal editors (e.g., the *Environmental DNA* journal), and experts in biodiversity and ‘omics data science’.
2. Identification of data components and formats suited for FAIR eDNA practice for both targeted and metabarcoding detection approaches.
3. Establishment of a comprehensive, FAIR eDNA (FAIRe) metadata checklist, with corresponding guidelines by adapting domain-specific usage of existing terms and adding new terms to existing standards (i.e., DwC, MIXS, DNA-derived extension).

# Findable, Accessible, Interoperable & Reusable eDNA data



**FIGURE 1** | Infographic of workflows and outcomes of FAIR eDNA practices.

4. Development of a customisable metadata template generator and an error-checking validation tool.

Together, these efforts aim to guide data providers and facilitate the unambiguous description of eDNA datasets that are aligned with FAIR data practices (Figure 1). Our goal is to provide a FAIR metadata checklist and guidelines as resources for existing data standards, infrastructures, and databases, helping them to address gaps to better meet their own objectives. The guidelines will also benefit individual eDNA science practitioners and labs by improving data management standardization, making their data more FAIR not only for reusers but also for internal organization.

Finally, we discuss additional steps to further advance FAIR practices within the eDNA community and maximize the utility of eDNA data. The proposed FAIR guidelines will evolve based on user feedback and the emergence of new methods and technologies, with the most up-to-date version available at <https://fair-edna.github.io/index.html>.

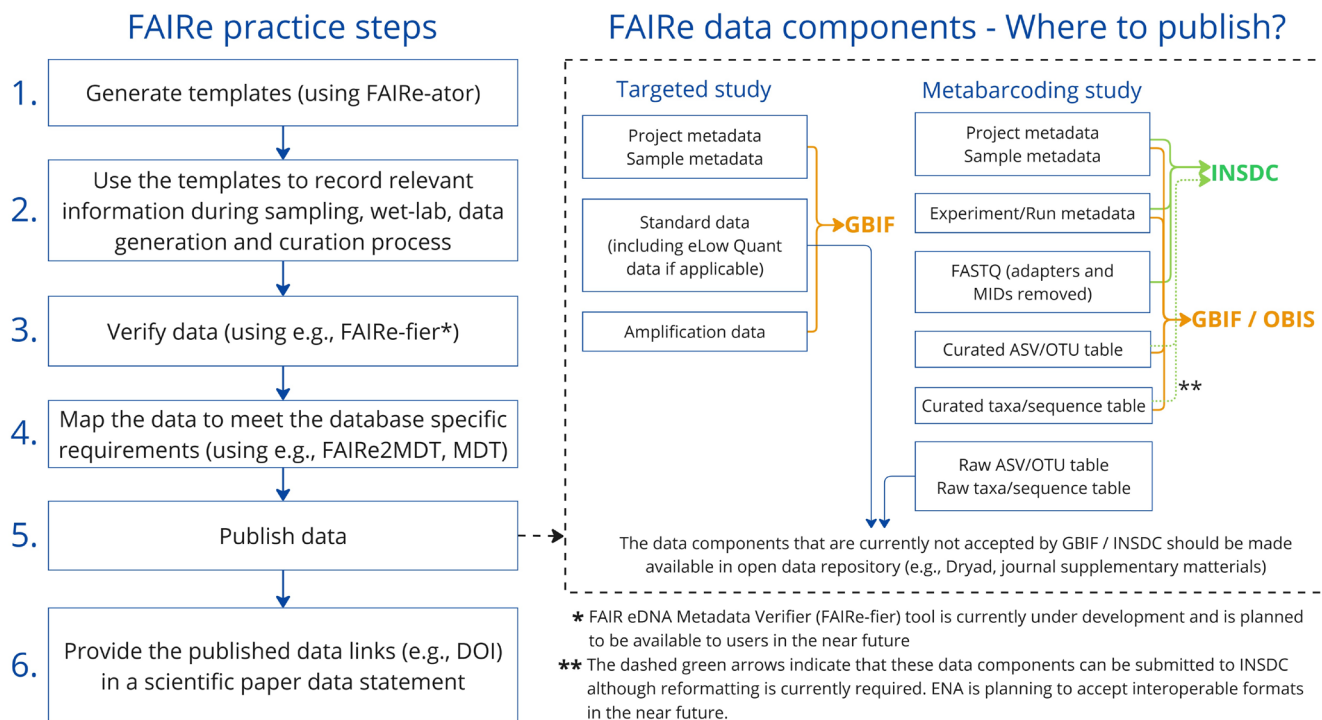
The proposed FAIR guidelines currently focus on the two major eDNA detection approaches: targeted (e.g., qPCR) and metabarcoding assay detections. While we acknowledge the rapid development of diverse techniques, such as multiplex and

high-throughput PCR and DNA metagenomics, these two approaches encompass the major eDNA workflows (Takahashi et al. 2023).

We also recognize that there are critical aspects of FAIR eDNA data that are not covered here. For example, the importance of FAIR code for reproducibility in bioinformatics and analysis workflows, as well as for enhancing the FAIRness of data supporting any science, is not covered here, but it has been addressed in recent studies by Ivimey-Cook et al. (2023), Thrall et al. (2023), Thalinger et al. (2021), Cadwallader et al. (2022) and Jenkins et al. (2023). This paper does not take a position on access and benefit-sharing (ABS) issues related to Digital Sequence Information (DSI) (<https://www.cbd.int/dsi-gr>). While ABS concerns often focus on the use of publicly available DSI, considerations around how DSI is shared also play a role in compliance with international, regional, and national regulations. Any data management plan can carefully account for these rules to ensure proper adherence.

## 4 | FAIR eDNA (FAIR) Practice Guidelines

In this section, we outline the various data components and formats, present a comprehensive FAIR eDNA (FAIR) Metadata



**FIGURE 2** | FAIR eDNA (FAIRe) practice flowchart (left) and the links between each of the data components and databases to publish (right).

Checklist, and provide example datasets, along with scripts and tools to guide users through the data formatting process. We also introduce existing databases where eDNA data can be shared and made publicly accessible. Figure 2 outlines the overall FAIRe practice steps, and the below sections contain the details of each step. The current FAIRe frameworks were developed using the following resources as foundational starting points.

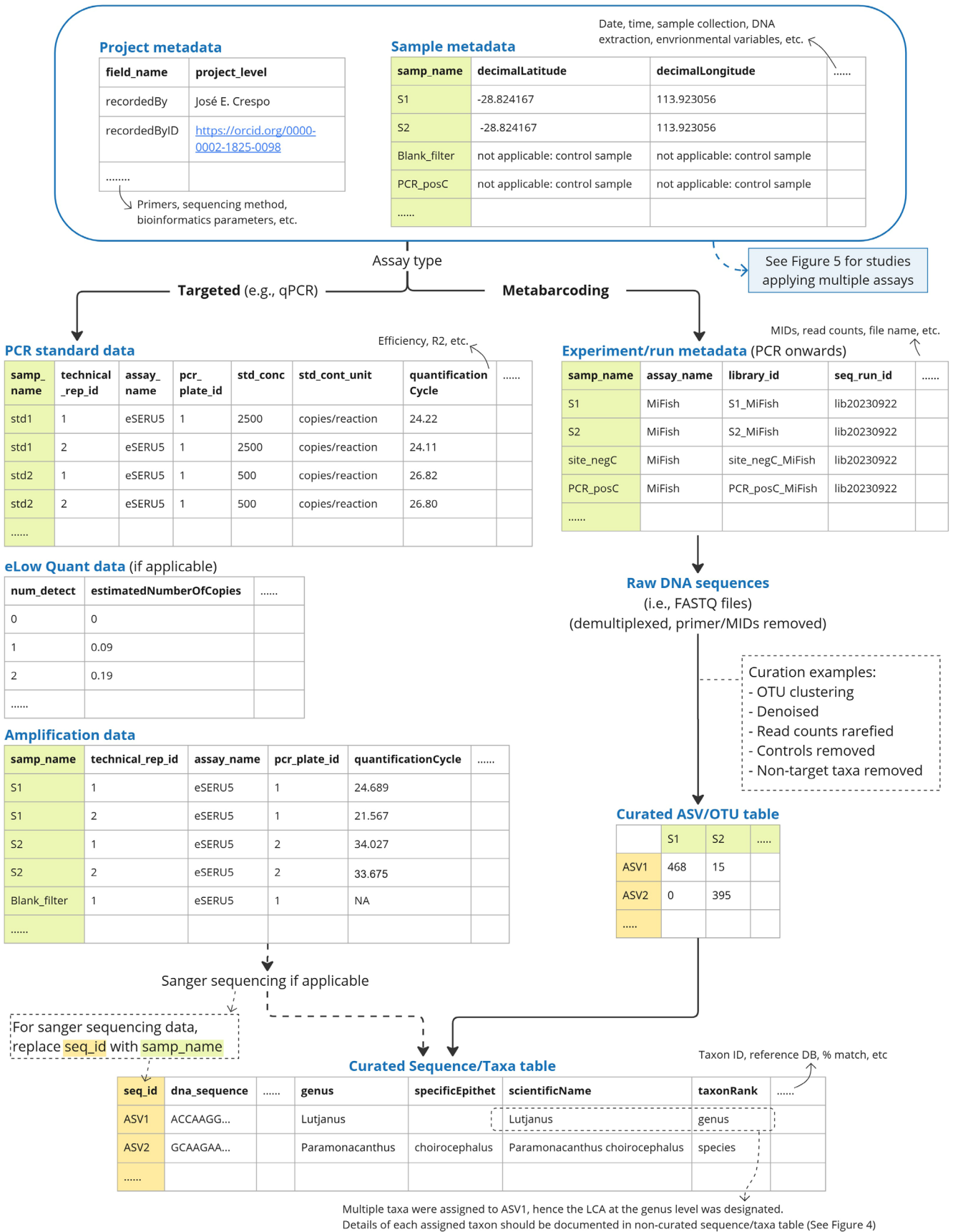
- Darwin Core Quick Reference Guide. <https://dwc.tdwg.org/terms/>.
- Minimum Information about any (x) Sequence (MIxS) standard. <https://genomicsstandardsconsortium.github.io/mixs/>.
- Darwin Core extension of DNA derived data. [https://rs.gbif.org/extension/gbif/1.0/dna\\_derived\\_data\\_2024-07-11.xml](https://rs.gbif.org/extension/gbif/1.0/dna_derived_data_2024-07-11.xml).
- Abarenkov et al. (2023) Publishing DNA-derived data through biodiversity data platforms, v1.3. Copenhagen: GBIF Secretariat. <https://doi.org/10.35035/doc-vf1a-nr22>.
- The OBIS manual. <https://manual.obis.org/GBIF> Metabarcoding Data Toolkit. <https://mdt.gbif.org/>.

#### 4.1 | Data Formats and Components

The list below and Figures 3–5 outline the associated eDNA data components and their contents and purposes. In our proposed data structure (Figure 3), all data components, except DNA sequence data (i.e., raw data in FASTQ and, optionally, representative sequences in FASTA), are organized in a tabular format. They can be saved as individual worksheets in spreadsheet workbooks, as character-separated values text files (tab

or comma), or as a combination of these (refer to various example datasets in <https://github.com/FAIR-eDNA/FAIR-eDNA.github.io/tree/main/docs/examples>). This allows data to be conceptually divided into separate modules that can be linked using unique identifiers, such as `project_id`, `sample_name`, and `assay_name`, as well as to accommodate users' needs and data wrangling preferences.

- Project metadata (projectMetadata)
  - Contents: Information that applies to an entire project and dataset. This includes a project name, reference associated with the datasets from which the datasets were derived (e.g., bibliographic references to studies, DOI of published, and associated data), and metadata describing workflows including PCR, sequencing, and bioinformatic steps.
  - Purpose: To describe project information and methodologies that apply to all samples, thereby avoiding the need to propagate identical entries to all samples in sample metadata. Information, such as bioinformatic quality filtering parameters and thresholds, enables data reusers to assess whether the data quality meets the specific requirements for certain types of reuse. Standardization of such information is particularly important as minimum requirements for amplification analysis, bioinformatic workflows, and quality assurance levels vary between studies and applications, including those reusing shared data.
  - Note: Multiple primer sets are often applied within one project to target more than one taxon. In this case, practitioners would follow the format in Figure 5 to record assay-specific workflows and outputs.
- Sample metadata (sampleMetadata)



**FIGURE 3** | Data components and their recommended structures that promote FAIR eDNA.

Non-curated ASV/OTU table → Control samples

	S1	S2	site_negC	PCR_posC	.....
ASV1	468	15	0	0	
ASV2	13	395	15	0	
ASV3	1	0	0	9852	
.....					

→ ASV of a positive control, which has been omitted in a curated ASV/OTU table

Non-curated Sequence/Taxa table

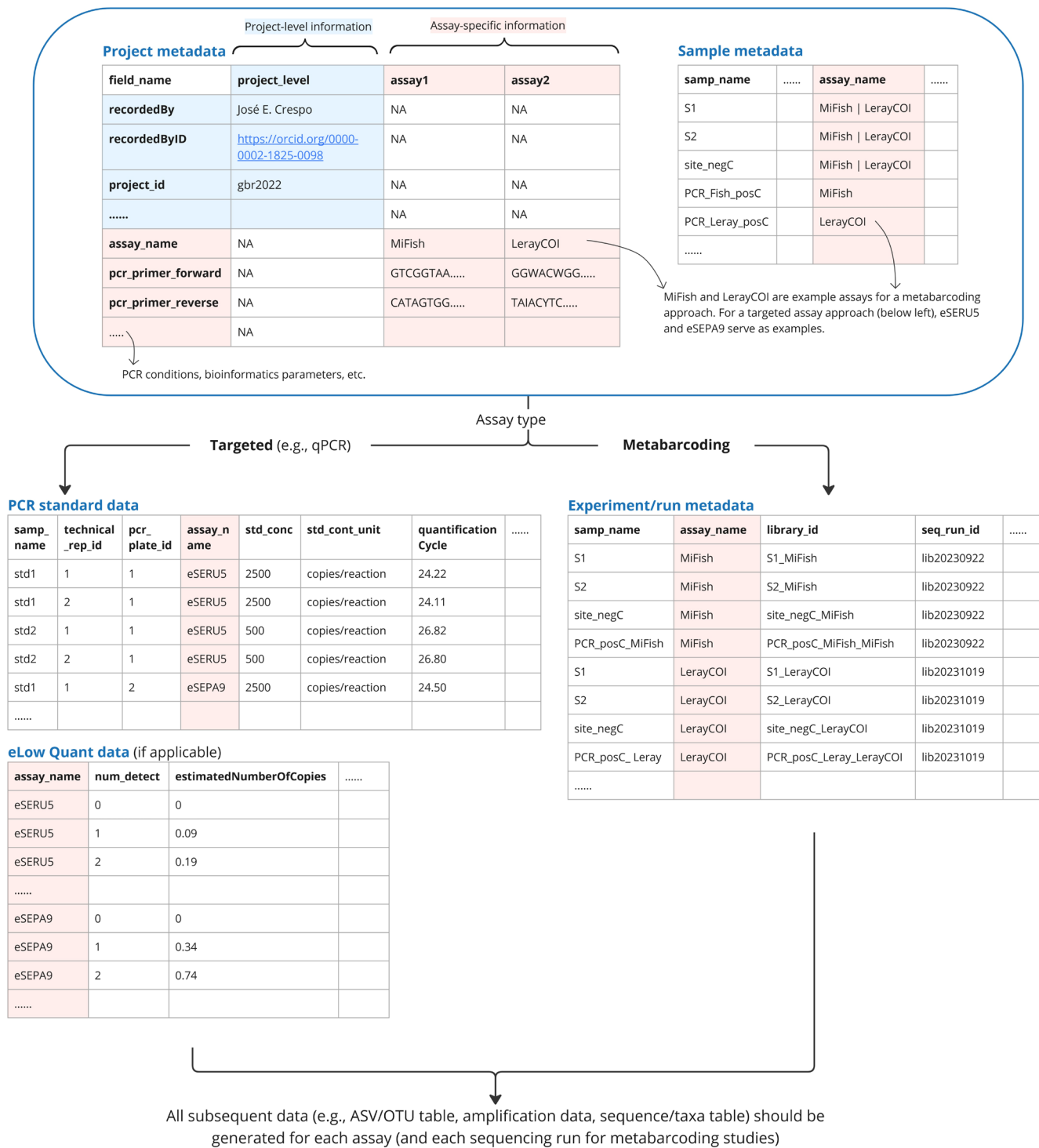
seq_id	dna_sequence	.....	genus	specificEpithet	scientificName	taxonRank	taxonID	taxonID_db	ref_db	accession_id	percent_match	percent_query_cover
ASV1	ACCAAGG...		Lutjanus	malabaricus	Lutjanus	genus	218797	NCBI	NCBI GenBank	MT123007.1	100	100
ASV1	ACCAAGG...		Lutjanus	erythropterus	Lutjanus	genus	211835	NCBI	NCBI GenBank	NC_031331.1	99.5	100
ASV2	GCAAGAA...		Paramonacanthus	choirocephalus	Paramonacanthus choirocephalus	species	392923	NCBI	NCBI GenBank	MN473743.1	97.537	100
ASV3	ACAAGATT...		Danio	rerio	Danio rerio	species	7955	NCBI	NCBI GenBank			
.....												

→ Including OTUs/ASVs that have been removed from curated ASV/OTU tables (e.g., positive control sequences, unassigned sequences).

→ Multiple taxa assigned to ASV1. Note only the lowest common ancestor (LCA) is reported in the curated Sequence/Taxa table.

**FIGURE 4** | Optional intermediate file formats for metabarcoding outputs.

- Contents: Information for each sample (where this cannot be stated at the project level in project metadata), including sampling date, location, methodologies (e.g., sample collection, preservation, storage, filtration, and DNA extraction), and environmental variables (i.e., temperature, water depth).
- Purpose: To record sample-specific information through the workflows from sampling to bioinformatics.
- Note: Including control samples in sample metadata can enhance data transparency and FAIRness. The metadata checklist includes terms such as `samp_category`, `neg_cont_type`, `pos_cont_type`, and `rel_cont_id` to describe control types and their links to eDNA samples.
- Experiment/run metadata (experimentRunMetadata)
  - Contents: Sample-specific information for PCR, library preparation (experiment) and sequencing (run) workflows.
  - Purpose: To record workflows, from PCR onward, that are specific to each sample and assay.
  - Note: Each entry (row) has a unique library ID (`lib_id`) and multiplexing identifiers (MIDs) (`mid_forward`, `mid_reverse`).
- PCR standard data (stdData)
  - Contents: Detailed information of PCR standards, including starting input quantity, Ct/Cq values, and standard curve parameters (e.g., efficiency,  $R^2$ ).
  - Purpose: To improve reproducibility and allow data reusers to regenerate the estimated copy numbers using their own approach. The entries in `assay_name` and `pcr_plate_id` in the PCR standard data and amplification data allow linking of the standards and eDNA samples.
- eLow Quant data (eLowQuantData)
  - Contents: Outputs from the eLow Quant binomial approach for PCR standards if applicable (refer to Lesperance et al. 2021).
  - Purpose: To improve reproducibility and allow data reusers to regenerate the estimated copy numbers using their own approach.
- Targeted assay amplification data (ampData)
  - Contents: Raw amplification data of each PCR technical replicate performed on eDNA samples, as well as estimated copy numbers based on standard curves (if applicable) and detection status of targeted taxon in each biological replicate (field sample).
  - Purpose: The Ct/Cq values of each technical replicate, combined with the standard information above, enable the re-conversion of the raw amplification data to concentration or detection status in future studies. Detect/non-detect records of a target taxon at a biological replicate level allow prompt reuses of taxon occurrence records in future studies, combined with sample metadata. They also standardize eDNA data in the context of biodiversity data produced using other methods and consequently make the collective data and underlying research discoverable and reusable for wider audiences.
  - Note: “NA” (not zero) should be used under the term *quantificationCycle* (i.e., Ct/Cq values) to indicate no amplification occurred, because zero could be interpreted as an extremely high DNA concentration, which may or may not be target DNA that would be unquantifiable due to being outside the standard curve range.
- Raw DNA sequences



**FIGURE 5** | Metadata formats for a project applying multiple assays.

- Contents: DNA sequences in FASTQ format, demultiplexed for each sample. The sequences of primers, adapters, and MIDIs must be removed.
- Purpose: To allow data reusers to repeat quality filtering, denoising, and taxonomic assignment using their own algorithms and thresholds to fit their purposes. To enable re-analyses using different or improved pipelines/algorithms. To enhance the reproducibility of studies.
- Non-curated ASV/OTU table (otuRaw) (Optional) (Figure 4)
  - Contents: The initial ASV/OTU table produced in a bioinformatic pipeline, containing absolute sequence read counts of all ASV/OTUs (including contaminant/nontargeted taxa and unassigned ASV/OTUs) across all samples (including controls). Each ASV/OTU will be provided with a unique seq\_id at this stage.
  - Purpose: To communicate objective information of read counts of all OTUs/ASVs in all samples, and

allow data reusers to modify or re-run the data curation steps if required. The inclusion of control samples, suspected contaminants, and nontargeted taxa allows the investigation of common contamination sequences, taxa, and sources across multiple studies, and aids in potentially identifying how to minimize them.

- Note: The column names in ASV/OTU tables should be either `samp_name` or `lib_id`, making sure they maintain clear links between the experiment/run metadata and the ASV/OTU table.
- Note: Although it is called a “raw” ASV/OTU table, some level of compression, quality filtering, and curation may already have been applied, such as filtering based on minimum length, read counts, error rates, and OTU clustering. The extent of data curation performed on the non-curated ASV/OTU table can vary between studies depending on the bioinformatics pipeline used, and this should be described under the term `otu_raw_description` in the project metadata.
- Curated ASV/OTU table (`otuFinal`)
  - Contents: An ASV/OTU table resulting from various data curation processes, including denoising (i.e., custom or LULU curation), removing control samples, and eliminating suspected contamination and nontarget taxa. This table should ideally be identical to, or correspond closely to, the one used for the final (e.g., ecological) analyses published in associated scientific papers.
  - Purpose: To facilitate reproducibility of original (ecological) analyses and general reuse of a curated/cleaned version of the data. To improve interoperability with other types of biodiversity data, and thereby facilitate incorporation in general biodiversity databases, and make the data and underlying research discoverable and reusable for wider audiences.
  - Note: The column names in ASV/OTU tables should be either `samp_name` or `lib_id`, making sure they maintain clear links between the experiment/run metadata and the ASV/OTU table.
  - Note: In some studies, researchers may choose to collapse multiple ASVs/OTUs assigned to the same taxon, summing their read counts. We, however, strongly encourage data providers to submit ASV/OTU tables in their original, non-collapsed form. Preserving each ASV/OTU with its individual read counts and sequence ensures better interoperability and prevents the loss of potentially valuable ASVs/OTUs due to uncertain or erroneous taxonomic annotations resulting from incomplete or imperfect reference databases at the time of analysis.
  - Note: Descriptions on how data were curated to produce the curated ASV/OTU table should be recorded under the term `otu_final_description` in the project metadata.
- Non-curated Sequence/Taxa table (`taxaRaw`) (Optional) (Figure 4)
  - Contents: In addition to the contents in the curated Sequence–Taxa table (below), this table includes ASVs/

OTUs that were excluded from the curated ASV/OTU table (e.g., nontarget taxa, contaminants, positive control sequences). It may also contain multiple rows for a single ASV/OTU assigned to multiple taxa, with information for each assigned taxon.

- Purpose: To enhance data transparency by providing access to ASVs, OTUs, and taxa excluded from the curated ASV/OTU table. To allow data reusers the opportunity to reassess and rerun specific data curation steps.
- Note: When assigning a lowest common ancestor (LCA) to a single ASV/OTU, multiple rows should be generated for the ASV/OTU to capture information about each of the assigned taxa. Users may apply a similarity threshold (e.g., 97%) and include all reference sequences within this range, or a representative subsample of them, in the output.
- Refer to the example in Figure 4 (ASV1 in the non-curated Sequence/Taxa table).
- Curated Sequence/Taxa table (`taxaFinal`)
  - Contents: DNA sequence of each `seq_id` listed in the curated ASV/OTU table. If a taxon was assigned, also include assigned taxonomy and assigned quality assurance parameters (e.g., % identity, % query coverage). If multiple taxa are assigned to a single ASV/OTU, the LCA should be used as the assigned taxon.
  - Purpose: To allow data reusers to assess the specificity and accuracy of the inferred taxonomy, and to perform taxonomic re-annotation using different or updated sequence reference databases.

## 4.2 | Consistent Identifiers and File Naming

To ensure machine readability and long-lasting reference to a digital resource, consistent, persistent sample and sequence identifiers (`samp_name`, `lib_id` and `seq_run_id`) across datasets (Damerow et al. 2021; McMurphy et al. 2017) must be used. Similarly, each file must have a clear, unambiguous name, consisting of `project_id`, `assay_name`, and `seq_run_id` (refer to Table 2 and example datasets) to maintain clear links between files. For example, a curated ASV/OTU table should be named as `otuFinal_<project_id>_<assay_name>_<seq_run_id>` (e.g., `otuFinal_gbr2022_MiFish_lib20230922.csv`). If multiple data components are stored within a single spreadsheet workbook, the file name should be the `project_id` (e.g., `gbr2022.xlsx`). In this case, each worksheet name should follow the format in Table 2 without the `project_id` as it already appears in the main file name (e.g., `otuFinal_MiFish_lib20230922`).

## 4.3 | Metadata Checklist

We developed a FAIR eDNA (FAIRe) metadata checklist, providing a comprehensive vocabulary for describing different data components and methodologies (Appendix S1). Thorough documentation promotes improved transparency and reproducibility of studies, as well as enabling the evaluation of data suitability for specific reuse cases. The FAIRe checklist consists of 337 data terms (38 mandatory, 51 highly

**TABLE 2** | Standardized names for each data type. In the examples, *project\_id* = “gbr2022”, *samp\_name* = “S1”, *assay\_name* = “eSERUS” and “MiFish” for targeted and metabarcoding assay approaches, respectively, *seq\_run\_id* = “run20230922”.

Data type	Name format	Example
Project metadata	projectMetadata_<project_id>	projectMetadata_gbr2022.csv
Sample metadata	sampleMetadata_<project_id>	sampleMetadata_gbr2022.csv
PCR standard data	stdData_<project_id>	stdData_gbr2022.csv
eLow Quant data	eLowQuantData_<project_id>	eLowQuantData_gbr2022.csv
Amplification data	ampData_<project_id>_<assay_name>	ampData_gbr2022_eSERUS.csv
Experiment/run metadata	experimentRunMetadata_<project_id>	experimentRunMetadata_gbr2022.csv
Non-curated ASV/OTU table	otuRaw_<project_id>_<assay_name>_<seq_run_id>	otuRaw_gbr2022_MiFish_run20230922.csv
Curated ASV/OTU table	otuFinal_<project_id>_<assay_name>_<seq_run_id>	otuFinal_gbr2022_MiFish_run20230922.csv
Non-curated Sequence/Taxa table	taxaRaw_<project_id>_<assay_name>_<seq_run_id>	taxaRaw_gbr2022_MiFish_run20230922.csv
Curated Sequence/Taxa table	taxaFinal_<project_id>_<assay_name>_<seq_run_id>	taxaFinal_gbr2022_MiFish_run20230922.csv

recommended, 128 recommended and 120 optional terms), organized into workflow sections (e.g., sample collection, PCR, bioinformatics). Most of the mandatory, highly recommended, and recommended information is generated naturally during the course of any project. As a result, data meeting the minimum reporting requirements can be readily made available and submitted to public repositories.

#### 4.3.1 | Vocabulary Basis

The FAIR metadata checklist consists of terms sourced from existing data standards such as MIXS, DwC, and the DNA-derived data extension to DwC. MIXS offers several checklists and extensions relevant for eDNA data, including Minimum Information about a Marker Sequence (MIMARKS), Minimum Information about Metagenome or Environmental Sequence (MIMS), and extensions for various environments (e.g., water, sediment, soil, air, host associated, and symbiont associated). These standards were reviewed and incorporated into the FAIR metadata checklist where applicable to ensure comprehensive coverage of eDNA data needs. We have also incorporated terms from MIQE guidelines (Bustin et al. 2009), the single species eDNA assay development and validation checklist (Thalinger et al. 2021) and Minimum Information for eDNA and eRNA Metabarcoding (MIEM) guidelines (Klymus et al. 2024). The source and Uniform Resource Identifier (URI) for each term from existing standards are documented in the FAIR metadata checklist under the columns “source” and “URI”. Where necessary, modifications have been made to term names, descriptions, and examples for easier interpretation by the eDNA community. A total of 158 new terms were proposed to accommodate some diverse attributes of eDNA procedures and datasets, and to improve the possibility of evaluating fitness for reuse. Among these are

terms related to targeted assay detection workflows (e.g., *lod\_method*, *std\_seq*), bioinformatic tools, filtering parameters and cutoffs (e.g., *demux\_tool*, *demux\_max\_mismatch*), taxonomic assignment metrics (e.g., *percent\_match*), and taxon screening methods (i.e., *screen\_contam\_method*, *screen\_nontarget\_method*). The FAIR checklist also includes a range of environmental variables relevant to eDNA samples, sourced from the MIXS sample extensions including Water, Soil, Sediment, Air, HostAssociated, MicrobialMatBiofilm, and SymbiontAssociated. These variables are typically recorded during sampling events, and sharing them could enable more advanced modeling approaches among other applications.

If suitable term names are not available in the FAIR checklist, users should search for them in existing standards, such as MIXS (<https://genomicsstandardsconsortium.github.io/mixs>) and DwC (<https://dwc.tdwg.org/terms>), and use these standardized terms where possible. If relevant terms cannot be found in these resources, users may add new terms using clear, concise, and descriptive names within related tables.

#### 4.3.2 | Specified Use of Terms and Controlled Vocabularies

Several approaches were implemented to increase standardization of terms. Firstly, for some free-text terms from existing standards, we propose using controlled vocabularies. For example, the value for the *target\_gene* term should be selected from a list of 28 gene regions used for barcoding. This standardization prevents variations of COI such as CO1, COX1, COX1, Cytochrome oxidase I gene, Cytochrome c oxidase I, and spelling mistakes/variants of these from being entered. If a suitable value cannot be found among the specified values

of the vocabulary, it is possible to enter “other:” followed by a free-text description. Secondly, units of numeric variables are strictly specified based on the International System of Units wherever possible, and only numeric entries without units are allowed. When units are necessary for unambiguous communication, separate terms for measurement value and measurement unit are provided, with the unit restricted to a controlled vocabulary (e.g., `samp_size` for the numeric value and `samp_size_unit` with options of mL, L, mg, g, kg, cm<sup>2</sup>, m<sup>2</sup>, cm<sup>3</sup>, m<sup>3</sup>, and other). Thirdly, formats for some data terms are restricted, generally following other standards (i.e., DwC, MlXs) and data infrastructures (i.e., GBIF, OBIS, INSDC). For example, *decimalLatitude* and *decimalLongitude* are required terms and must be in decimal degrees using WGS84 datum. If records in an original datasheet follow other formats (e.g., degree minute second, UTM, datums other than WGS84), they should be stored under the terms *verbatimLatitude* and *verbatimLongitude* while *decimalLatitude* and *decimalLongitude* are also required to be entered. Similarly, sampling date and time are summarized under the term *eventDate*.

Which utilizes the ISO 8601 format (<https://www.iso.org/iso-8601-date-and-time-format.html>), followed by the difference from UTC time (e.g., “2008-01-23T19:23-06:00” in the time zone 6 h earlier than UTC). Original records of date and time before conversion into the ISO 8601 format should be stored under the terms *verbatimDate* and *verbatimTime*.

#### 4.3.3 | Missing Values

Information may be missing for various reasons, and historically, science has used a variety of conventions to denote missing data. Furthermore, geographical coordinates can be generalized or withheld to protect endangered species or culturally important sites for Indigenous People and Local Communities (Chapman 2020; Chapman and Wiczorek 2020; Frank et al. 2015; <https://fnigc.ca>). The reasons for withholding or generalizing data should be described under the terms *informationWithheld* and *dataGeneralization*, respectively, in project metadata. When a value is missing from a mandatory term, it is required to provide the reason using the INSDC missing value controlled vocabulary (<https://www.insdc.org/submitting-standards/missing-value-reporting>). It is also recommended to apply this approach for non-mandatory terms. Below, we provide the complete list of missing value terms, including their full hierarchical structure for clarity. Only the bold texts should be used to describe the missing values (i.e., do not include the parenthetical explanations or add your own reasons).

- **Not applicable: control sample**
- **Not applicable: sample group**
- **Not applicable** (other reasons not applicable)
- **Missing: not collected: synthetic construct**
- **Missing: not collected: lab stock**
- **Missing: not collected: third party data**
- **Missing: not collected** (other reasons not collected)

- **Missing: not provided: data agreement established pre-2023**
- **Missing: not provided** (other reasons not provided)
- **Missing: restricted access: endangered species**
- **Missing: restricted access: human-identifiable**
- **Missing: restricted access** (other reasons restricted access)

#### 4.4 | Data Submission/Publication

Publishing eDNA datasets to databases such as GBIF, OBIS, and INSDC is an essential step in ensuring FAIR data practices. These platforms ensure data findability and accessibility for scientific research and decision making that rely on open data, either through application programming interfaces (APIs) or sophisticated web-browser interfaces. They offer data validation procedures during submission and provide additional standardization to ensure interoperability. Persistent sample and sequence identifiers are provided upon the submission of data to these databases, which are essential for machine-readability and long-lasting reference to the data. For example, in GBIF, dataset authors and publishers receive credit through unique dataset Digital Object Identifiers (DOIs) that enable citation tracking and automatic usage reporting.

Because not all eDNA data components can be sufficiently accommodated in any one of these databases, they should be submitted in parts to the most suitable databases (Figure 2). These are some current practices for sharing FAIR eDNA data through currently established repositories/databases/infrastructures:

1. Submit raw DNA sequence data and metadata to nucleotide databases such as INSDC (ENA, NCBI, or DDBJ). For instance, NCBI allows submission of project metadata via BioProject, sample metadata via BioSample, and raw sequences via the Sequence Read Archive (SRA) (Barrett et al. 2012). This links metadata with sequence data, enabling users to retrieve complete datasets by querying BioProject or BioSample records.
2. Publish the derived biodiversity data—that is, inferred taxon occurrences with sequences and metadata—to biodiversity databases such as GBIF or GBIF and OBIS (if marine data). Both GBIF and OBIS rely on data being formatted according to the DwC standard with the DNA-derived data extension as described in the co-authored guide (Abarenkov et al. 2023). The dataset is converted into so-called Darwin Core Archive (DwC-A) that can be indexed by GBIF and OBIS.
3. Archive the remaining data components (i.e., standard data for targeted assay studies and raw ASV/OTU table for metabarcoding studies), formatted according to the FAIR guidelines, in open data repositories such as Dryad, Zenodo, and Figshare or as Supporting Information in journals. This approach ensures comprehensive storage of raw data that are not addressed in the above steps. However, datasets in these repositories are not indexed to

the extent required for effective searchability and interoperability within larger eDNA, biodiversity, or nucleotide data platforms.

4. Publishing protocols and sharing bioinformatics and analytical codes (Jenkins et al. 2023; Samuel et al. 2021; Teytelman et al. 2016) through open-source platforms such as protocols.io (<https://www.protocols.io>), WorkflowHub (<https://workflowhub.eu>), and Better Biomolecular Ocean Practices (BeBOP) (<https://github.com/BeBOP-OBON>) can further support FAIR principles.
5. The linkage between all the above data, protocols, and codes can be strengthened by assigning DOIs and documenting them in data accessibility statements within journal articles, as well as in the metadata terms such as `seq_archive`, `code_repo`, and `associated_resource`.

The guidelines provided here result in data formatted slightly different from those accepted by nucleotide and biodiversity databases. Hence, minor reformatting is required. One example is the necessary data format when multiple assays were applied in a single project. GBIF recently launched the Metabarcoding Data Toolkit (MDT), a user-friendly web application (<https://www.gbif.org/metabarcoding>) that reshapes tabular metabarcoding data—similar to the format proposed here—and publishes it to GBIF and OBIS complying with the guidelines in Abarenkov et al. (2023). The GBIF MDT currently requires a separate input dataset for each assay (hence information from project and sample metadata are repeated), whereas the FAIR checklist allows the metadata from multiple assays to be combined (Figure 5). An R-script (FAIRe2MDT) has been developed (R Core Team 2024) to convert FAIRe data templates for GBIF and OBIS publication via MDT, bridging the differences in data formats between FAIRe and GBIF standards (refer to section ‘Available example datasets, scripts and tools’). Using these tools in combination streamlines the publication of eDNA data to GBIF and OBIS. Overall, there is currently a strong effort by biodiversity data platforms to improve the suitability of depositing eDNA-based data.

#### 4.5 | Available Example Datasets, Scripts and Tools

To provide eDNA practitioners with suggestions for how to format and publish their data, various example datasets are available at <https://github.com/FAIR-eDNA/FAIR-eDNA.github.io/tree/main/docs/examples> (refer to Appendix S2 and S3 for the list of example datasets using targeted and metabarcoding assays, respectively). The below scripts and tools were developed to aid data formatting according to the guidelines and are made available at <https://github.com/orgs/FAIR-eDNA/repositories>.

- FAIRe-ator (FAIR eDNA template generator): This R function creates data templates based on user-specified parameters, such as assay type (i.e., targeted or metabarcoding assay), sample type (e.g., water, sediment), and the number of assays applied. Additionally, the function allows users to input project ID and assay name(s), which ensures correct file name formatting and pre-fills the `project_id`

and `assay_name` terms in the template. While the full template (Appendix S4) includes data components and metadata terms for all assay and sample types, FAIRe-ator produces study-specific templates tailored to individual project needs.

- FAIRe-fier (FAIR eDNA metadata verifier): This planned tool has a working prototype already in place. The plan is to make this metadata verifier publicly accessible through a web interface, allowing users to upload their metadata and validate it without requiring any scripting knowledge. The tool checks various components, including whether mandatory terms are completed or, if not, whether a valid reason has been provided following the missing value guidelines. It also verifies that controlled vocabulary entries and fixed-format terms are correctly formatted. If any issues are found, users receive output with warning and error messages, indicating where corrections are needed. Formatted output will be produced when there is no warning message, allowing users to submit it to the respective publishers.
- FAIRe2MDT: This R script converts FAIR eDNA data templates for GBIF submission via MDT.

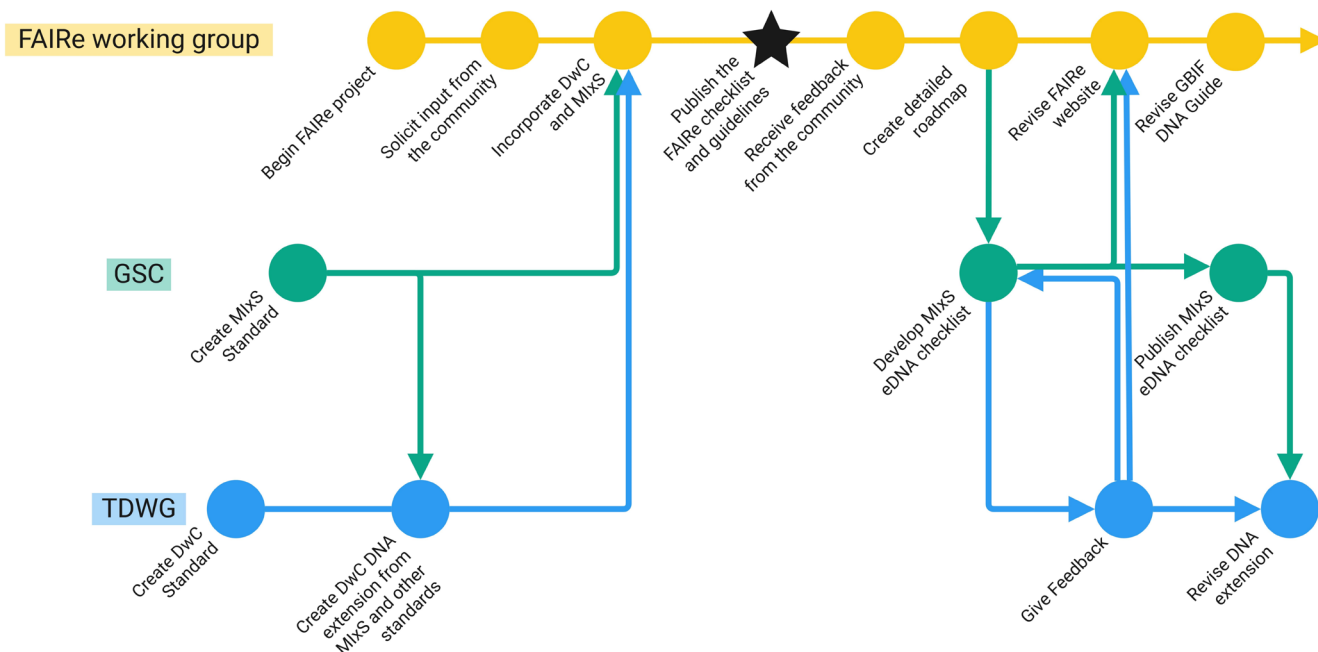
## 5 | Planned Next Steps

The current guidelines can substantially empower users to enhance the FAIRness of eDNA data, but other contributions will also make a difference.

### 5.1 | Raising Awareness

Our next step focuses on raising awareness of FAIR principles, existing guidelines, and available tools to promote their adoption with the eDNA communities. To achieve this, the lead author will host workshops with eDNA practitioners over the coming year, offering hands-on opportunities to introduce FAIR data principles and provide practical guidance on implementing standardized formatting protocols. The first workshop took place at the eDNA conference in Wellington, New Zealand, in February 2025. More than a training exercise, these workshops enable establishing direct communications, identifying communities’ needs, and addressing challenges to adopting FAIR data practices within their workflows. This step is particularly important as open and FAIR data can only be achieved through eDNA practitioners’ collaborative efforts and relies on their commitment to making data accessible for broader use.

Several studies have explored Open Science behaviors and identified the barriers among scientists (e.g., Norris and O’Connor 2019; Tenopir et al. 2015). These barriers include technical and resource limitations, such as lack of standards, tools, time and skills to navigate required data management systems (Tedersoo et al. 2021; Tenopir et al. 2015). Additionally, multiple perceived barriers exist, including unawareness of the value of data for others, fear of scrutiny due to potential mistakes, resistance to openly sharing data given large efforts invested to secure funding in the face of limited resource for scientific research and resistance to changing existing practices (Norris and O’Connor 2019; Tenopir et al. 2015).



**FIGURE 6** | Roadmap for continued development of the FAIR eDNA (FAIRe) checklist and guidelines, with time represented from left to right and the star representing the current point in time.

To address these challenges, communicating the diverse benefits of data sharing beyond reproducibility and documentation is important. In addition to contributing to the public good, data sharing leads to various benefits to data providers. These include increased visibility and citation of associated work (Colavizza et al. 2020; Piwowar et al. 2007; Wood-Charlson et al. 2022), direct citation metrics of data accessed through databases (e.g., GBIF), and enhanced collaborations and co-authorships, which all lead to improved professional stature (Bethlehem et al. 2022; McKiernan et al. 2016; Piwowar and Vision 2013; Whitlock 2011). Many eDNA practitioners, however, may not yet be fully aware of these benefits. Establishing open and direct communications with them through workshops will help us identify their unique barriers and needs, work collaboratively to overcome them, and foster a culture of FAIR data practices in the community.

## 5.2 | Revising, Updating and Integrating the Guidelines

As eDNA science is relatively new and rapidly evolving, new methods and technologies will emerge. To improve and adjust the FAIRe metadata checklist and formatting guidelines and keep them relevant, regular revisions and updates are essential. Input from data providers and reusers is vital to this process, and we welcome feedback at any time via <https://github.com/FAIR-eDNA/FAIR-eDNA.github.io/issues>. Updated guidelines will be available at <https://fair-edna.github.io>.

Leveraging the network of the co-authors (e.g., GSC, TDWG, ENA, GBIF, OBIS, and ALA), we aim to continuously revise and co-develop the FAIRe metadata checklist to integrate with established data standards and databases (Figure 6). Implementation and alignment of FAIRe procedures and formats will be supported through contributions to the existing guidelines for publishing

DNA-derived biodiversity data (Abarenkov et al. 2023), and development of scripts, like the FAIRe-ator and FAIRe2GBIF, and tools like the GBIF MDT. Integration of the FAIRe checklist into the GSC suite (i.e., MixS), that will facilitate its adoption within the INSDC system, will be pursued through collaboration with the GSC MixS Compliance and Interoperability Working Group (CIG) and with input from TDWG (Figure 6).

In the scope of existing agreements between TDWG and GSC to collaborate on the development of specifications for sequence-based biodiversity data, we plan to support the integration of the FAIRe checklist into the DwC DNA Derived Data extension. Continued revision, testing, and alignment with these organizations' objectives will help ensure successful implementation. Important milestones include the publication of updated guidelines for publishing DNA-derived biodiversity data (Abarenkov et al. 2023), the publication of the eDNA MixS checklist, and the publication of the FAIRe-aligned DwC DNA Derived Data extension. Achieving these milestones will serve as evidence that the FAIRe checklist is adequate and accepted by the eDNA community. Long-term maintenance is planned to be managed by these consortia and organizations, but will depend on sustained engagement and interest from the eDNA science community in the FAIRe checklist.

We have developed the current guidelines with great interest to integrate them into journal publication requirements. We have collaborated with Wiley and *Environmental DNA* to develop a roadmap for integrating FAIR data principles into their publication procedures. This roadmap involves three key steps: (1) strongly recommending adherence to the FAIRe guidelines upon journal article publication, (2) making compliance mandatory in due course, and (3) extending these requirements to additional journals including *Molecular Ecology* and *Ecology and Evolution*. While most journals mandate data sharing, submitting data in a FAIR manner is not yet a widespread practice

(Roche et al. 2015). Implementation of the FAIR data guidelines could be an effective strategy given the large number of eDNA studies published each year with the tremendous volume of genetic and species occurrence data they generate (Takahashi et al. 2023). [Correction added on 18 August 2025, after first on-line publication: *Molecular Ecology and Evolution* is corrected to *Molecular Ecology and Ecology and Evolution*.]

### 5.3 | Harmonizing FAIR and CARE

The CARE principles, which stand for Collective benefit, Authority to control, Responsibility, and Ethics, were developed by the Global Indigenous Data Alliance to protect the Indigenous rights and interests in Indigenous data, which include information, data, and traditional knowledge about their resources and environments (Carroll et al. 2021, 2020). The people- and purpose-oriented CARE principles are designed to complement the data-centric FAIR principles by ensuring that Indigenous data remain FAIR while centering Indigenous sovereignty (Mc Cartney et al. 2023). Implementing the CARE principles into FAIR data initiatives can ensure that data use respects Indigenous rights, serves a meaningful purpose, and promotes the wellbeing of Indigenous Peoples and Local Communities (IPLCs) (Carroll et al. 2021; O'Brien et al. 2024).

Mc Cartney et al. (2023) established a roadmap for the lifecycle of eukaryotic biodiversity sequencing data following the CARE principles. It identifies actions required from researchers throughout each study to build sustainable partnerships with IPLCs. In the context of FAIR initiatives, it is important to develop culturally aware linked metadata. For example, generalizing or withholding culturally sensitive data, such as geographical coordinates of culturally important sites, is respected and accepted, which can be mapped under the terms *dataGeneralization* and *information-Withheld* in the FAIR metadata checklist and other standards and databases (Chapman 2020). Information on access, use permission, and the person or organization owning and managing data rights, should be defined at the start of each project, and recorded under the terms *accessRights* and *rightsHolder*. Further works are needed in contextual metadata to document Traditional Knowledge Labels and Notices (Liggins et al. 2021; “Local Contexts – Grounding Indigenous Rights” 2023), provenance about place and people (Mc Cartney et al. 2023), and the cultural importance of species (Reyes-García et al. 2023). Meetings, workshops and collaboration across various stakeholders including IPLCs and eDNA communities can elucidate the benefits of eDNA data sharing in various aspects, exchange more information and raise discussions. All these efforts can harmonize FAIR and CARE, establish and retain the link between people and nature, and serve as a vehicle for knowledge transfer, capacity development, two-way science and biodiversity conservation.

## 6 | Conclusion

Our work advances the goal to make eDNA data FAIR by identifying essential data components and formats, developing an eDNA-specific, comprehensive metadata checklist that builds on existing biodiversity and molecular data standards, and providing data formatting guidelines for eDNA communities.

We also outline future steps that support effective integration of FAIR practices within these communities. Looking ahead, we are focused on creating tools to make FAIR procedures automated and hassle-free, ensuring that FAIR data practices become the norm among eDNA practitioners. This is an ongoing, community-driven project, continuously refined with input from broad communities to enhance and adapt the proposed guidelines presented here, which is necessary in the rapidly evolving eDNA field.

### Author Contributions

Conceptualization: M.T., O.B., M.L., G.J., T.G.F., and J.P. Validation: all authors. Data curation/visualization: B.D., C.V.R., K.S., K.W., L.R.H., L.R.T., M.L.D.L., M.T., N.A.P., N.D., C.C.H., and R.H. Software: M.T., N.D., S.F., and S.Y.Y., with feedback provided by K.S., K.W., L.R.H., L.R.T., M.L.D.L., N.A.P., R.H., S.P.J., T.S.J., and T.G.F. Writing – Original Draft: M.T. Writing – Review and Editing: all authors. Project administration: M.T. Funding acquisition: O.B.

### Acknowledgments

We deeply and gratefully acknowledge the invaluable contributions of Prof. Louis Bernatchez, who passed away in 2023. Prof. Bernatchez's enthusiasm for the topic, plus his extensive networks, were pivotal in establishing the FAIR working group. His commitment to the fields of molecular ecology, conservation, and eDNA, as well as his dedication to integrating FAIR guidelines into journal data-sharing policies, laid the foundation for this initiative. We extend our thanks to all participants of the eDNA biodiversity data workshop, co-hosted by GBIF and CSIRO in Canberra in October 2023. The workshop facilitated vibrant discussions on the opportunities and challenges in advancing FAIR eDNA practices, fostering collaboration across the global eDNA and biodiversity database communities. Special thanks go to Jodie van de Kamp, Nerida Wilson, Haylea Power, Chloe Anderson, and Alyssa Budd for their technical expertise and valuable advice on eDNA, taxonomy, and database development. We acknowledge Ooid Scientific (<https://ooidscientific.com/>) for their design of the infographic (Figure 1). This project was funded by the Environomics Future Science Platforms of the Commonwealth Scientific and Industrial Research Organization (CSIRO). Additionally, the contributions of T.G.F. and T.S.J. to this project have received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101057437 (BioDT project, <https://doi.org/10.3030/101057437>). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them. L.R.T. was supported by award NA21OAR4320190 to the Northern Gulf Institute from NOAA's Office of Oceanic and Atmospheric Research, U.S. Department of Commerce.

### Disclosure

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

### Conflicts of Interest

The authors declare no conflicts of interest.

### Data Availability Statement

The data and resources (including the FAIR metadata checklist, templates, example datasets, codes) derived from this article are available in the [Supporting Information](#) of this article and the GitHub repositories (<https://github.com/orgs/FAIR-eDNA/repositories>). All the above resources are also accessible via <https://fair-edna.github.io/index.html>.

## References

- Abarenkov, K., A. Andersson, A. Bissett, et al. 2023. "Publishing DNA-Derived Data Through Biodiversity Data Platforms."
- Abbott, C., M. Coulson, N. Gagné, et al. 2021. *Guidance on the Use of Targeted Environmental DNA (eDNA) Analysis for the Management of Aquatic Invasive Species and Species at Risk*. Canadian Science Advisory Secretariat (CSAS).
- Arita, M., I. Karsch-Mizrachi, and G. Cochrane. 2021. "The International Nucleotide Sequence Database Collaboration." *Nucleic Acids Research* 49: D121–D124. <https://doi.org/10.1093/nar/gkaa967>.
- Armit, C., M. A. Tuli, and C. I. Hunter. 2022. "A Decade of GigaScience: GigaDB and the Open Data Movement." *GigaScience* 11: giac053. <https://doi.org/10.1093/gigascience/giac053>.
- Australian Research Council. 2021. <https://www.arc.gov.au/>.
- Baillie, S. M., C. McGowan, S. May-McNally, R. Leggett, B. J. Sutherland, and S. Robinson. 2019. *Environmental DNA and Its Applications to Fisheries and Oceans Canada: National Needs and Priorities*. Fisheries and Oceans Canada.
- Barrett, T., K. Clark, R. Gevorgyan, et al. 2012. "BioProject and BioSample Databases at NCBI: Facilitating Capture and Organization of Metadata." *Nucleic Acids Research* 40: D57–D63. <https://doi.org/10.1093/nar/gkr1163>.
- Beck, M. W., C. O'Hara, J. S. Stewart Lowndes, et al. 2020. "The Importance of Open Science for Biological Assessment of Aquatic Environments." *PeerJ* 8: e9539. <https://doi.org/10.7717/peerj.9539>.
- Berry, O., S. Jarman, A. Bissett, et al. 2021. "Making Environmental DNA (eDNA) Biodiversity Records Globally Accessible." *Environmental DNA* 3: 699–705. <https://doi.org/10.1002/edn3.173>.
- Bethlehem, R. A., J. Seidlitz, S. R. White, et al. 2022. "Brain Charts for the Human Lifespan." *Nature* 604: 525–533.
- Blancher, P., E. Lefrançois, F. Rimet, et al. 2022. "A Strategy for Successful Integration of DNA-Based Methods in Aquatic Monitoring." *Metabarcoding and Metagenomics* 6: 258.
- Borisenko, A., R. Young, and R. Hanner. 2024. "A Lab-Centric, Workflow-Based Data Management System for Environmental DNA Research." *Research Ideas & Outcomes* 10: e120483.
- Bustin, S. A., V. Benes, J. A. Garson, et al. 2009. "The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments." *Clinical Chemistry* 55: 611–622. <https://doi.org/10.1373/clinchem.2008.112797>.
- Cadwallader, L., F. Mac Gabhann, J. Papin, and V. E. Pitzer. 2022. "Advancing Code Sharing in the Computational Biology Community." *PLoS Computational Biology* 18: e1010193. <https://doi.org/10.1371/journal.pcbi.1010193>.
- Carroll, S. R., I. Garba, O. L. Figueroa-Rodríguez, et al. 2020. "The CARE Principles for Indigenous Data Governance." *Data Science Journal* 19: 43. <https://doi.org/10.5334/dsj-2020-043>.
- Carroll, S. R., E. Herczog, M. Hudson, K. Russell, and S. Stall. 2021. "Operationalizing the CARE and FAIR Principles for Indigenous Data Futures." *Scientific Data* 8: 108. <https://doi.org/10.1038/s41597-021-00892-0>.
- Ceballos, G., P. R. Ehrlich, A. D. Barnosky, A. García, R. M. Pringle, and T. M. Palmer. 2015. "Accelerated Modern Human-Induced Species Losses: Entering the Sixth Mass Extinction." *Science Advances* 1: e1400253. <https://doi.org/10.1126/sciadv.1400253>.
- Chapman, A. 2020. "Current Best Practices for Generalizing Sensitive Species Occurrence Data."
- Chapman, A., and J. Wicczorek. 2020. "Georeferencing Best Practices."
- Coissac, E., T. Riaz, and N. Puillandre. 2012. "Bioinformatic Challenges for DNA Metabarcoding of Plants and Animals." *Molecular Ecology* 21: 1834–1847. <https://doi.org/10.1111/j.1365-294X.2012.05550.x>.
- Colavizza, G., I. Hrynaszkiewicz, I. Staden, K. Whitaker, and B. McGillivray. 2020. "The Citation Advantage of Linking Publications to Research Data." *PLoS One* 15: e0230416. <https://doi.org/10.1371/journal.pone.0230416>.
- Crandall, E. D., R. H. Toczydlowski, L. Liggins, et al. 2023. "Importance of Timely Metadata Curation to the Global Surveillance of Genetic Diversity." *Conservation Biology* 37: e14061. <https://doi.org/10.1111/cobi.14061>.
- Cross-Infrastructure Collaboration With ENA Improves Processing, Quality of DNA-Derived Occurrences. 2021. <https://www.gbif.org/news/3jpFT9gysj9zwcupx1gOKU/cross-infrastructure-collaboration-with-ena-improves-processing-quality-of-dna-derived-occurrences>.
- Damerow, J. E., C. Varadharajan, K. Boye, et al. 2021. "Sample Identifiers and Metadata to Support Data Management and Reuse in Multidisciplinary Ecosystem Sciences." *CODATA* 20: 11. <https://doi.org/10.5334/dsj-2021-011>.
- Delrieu-Trottin, E., J. T. Williams, D. Pitassy, et al. 2019. "A DNA Barcode Reference Library of French Polynesian Shore Fishes." *Scientific Data* 6: 114. <https://doi.org/10.1038/s41597-019-0123-5>.
- Diaz, S., J. Settele, E. Brondizio, et al. 2019. "Pervasive Human-Driven Decline of Life on Earth Points to the Need for Transformative Change." *Science (New York, N.Y.)* 366: eaax3100. <https://doi.org/10.1126/science.aax3100>.
- Ficetola, G. F., C. Miaud, F. Pompanon, and P. Taberlet. 2008. "Species Detection Using Environmental DNA From Water Samples." *Biology Letters* 4: 423–425. <https://doi.org/10.1098/rsbl.2008.0118>.
- Ficetola, G. F., J. Pansu, A. Bonin, et al. 2015. "Replication Levels, False Presences and the Estimation of the Presence/Absence From eDNA Metabarcoding Data." *Molecular Ecology Resources* 15: 543–556. <https://doi.org/10.1111/1755-0998.12338>.
- Frank, R. D., A. Kriesberg, E. Yakel, and I. M. Faniel. 2015. "Looting Hoards of Gold and Poaching Spotted Owls: Data Confidentiality Among Archaeologists & Zoologists." *Proceedings of the Association for Information Science and Technology* 52: 1–10. <https://doi.org/10.1002/pra2.2015.145052010037>.
- Froslev, T. G., R. Ejrnæs, A. J. Hansen, et al. 2023. "Treated Like Dirt: Robust Forensic and Ecological Inferences From Soil eDNA After Challenging Sample Storage." *Environmental DNA* 5: 158–174. <https://doi.org/10.1002/edn3.367>.
- Gagné, N., L. Bernatchez, D. Bright, et al. 2021. *Environmental DNA (eDNA) Reporting Requirements and Terminology*. Canadian Standards Association.
- Goldberg, C. S., K. M. Strickler, and D. S. Pilliod. 2015. "Moving Environmental DNA Methods From Concept to Practice for Monitoring Aquatic Macroorganisms." *Biological Conservation* 183: 1–3.
- Gomes, D. G. E., P. Pottier, R. Crystal-Ornelas, et al. 2022. "Why Don't We Share Data and Code? Perceived Barriers and Benefits to Public Archiving Practices." *Proceedings of the Royal Society B: Biological Sciences* 289: 20221113. <https://doi.org/10.1098/rspb.2022.1113>.
- Goodwin, K. D., M. Weise, C. M. Aiello, et al. 2024. *National Aquatic Environmental DNA Strategy*. White House Office of Science, Technology, and Policy (OSTP).
- Hampton, S. E., C. A. Strasser, J. J. Tewksbury, et al. 2013. "Big Data and the Future of Ecology." *Frontiers in Ecology and the Environment* 11: 156–162. <https://doi.org/10.1890/120103>.
- Hashemzadeh Segherloo, I., S. N. Tabatabaei, E. Abdolahi-Mousavi, et al. 2022. "eDNA Metabarcoding as a Means to Assess Distribution

- of Subterranean Fish Communities: Iranian Blind Cave Fishes as a Case Study." *Environmental DNA* 4: 402–416. <https://doi.org/10.1002/edn3.264>.
- Hassenrück, C., T. Poprick, V. Helfer, M. Molari, R. Meyer, and I. Kostadinov. 2021. "FAIR Enough? A Perspective on the Status of Nucleotide Sequence Data and Metadata on Public Archives." <https://doi.org/10.1101/2021.09.23.461561>.
- Hebert, P. D. N., A. Cywinska, S. L. Ball, and J. R. deWaard. 2003. "Biological Identifications Through DNA Barcodes." *Proceedings of the Biological Sciences* 270: 313–321. <https://doi.org/10.1098/rspb.2002.2218>.
- Heidorn, P. B. 2008. "Shedding Light on the Dark Data in the Long Tail of Science." *Library* 57: 280–299. <https://doi.org/10.1353/lib.0.0036>.
- Helbing, C. C., and J. Hobbs. 2019. *Environmental DNA Standardization Needs for Fish and Wildlife Population Assessments and Monitoring*. CSA Group.
- Herold, P. 2015. "Data Sharing Among Ecology, Evolution, and Natural Resources Scientists: An Analysis of Selected Publications." *Journal of Librarianship and Scholarly Communication* 3: 1244. <https://doi.org/10.7710/2162-3309.1244>.
- Hutchison, V. B., T. E. Burley, K. W. Blasch, et al. 2024. *U.S. Geological Survey Data Strategy 2023–33 (No. 1517), Circular*. U.S. Geological Survey.
- Ivimey-Cook, E. R., J. L. Pick, K. R. Bairos-Novak, et al. 2023. "Implementing Code Review in the Scientific Workflow: Insights From Ecology and Evolutionary Biology." *Journal of Evolutionary Biology* 36: 1347–1356. <https://doi.org/10.1111/jeb.14230>.
- Jacobsen, A., R. De Miranda Azevedo, N. Juty, et al. 2020. "FAIR Principles: Interpretations and Implementation Considerations." *Data Intelligence* 2: 10–29. [https://doi.org/10.1162/dint\\_r\\_00024](https://doi.org/10.1162/dint_r_00024).
- Jenkins, G. B., A. P. Beckerman, C. Bellard, et al. 2023. "Reproducibility in Ecology and Evolution: Minimum Standards for Data and Code." *Ecology and Evolution* 13: e9961. <https://doi.org/10.1002/ece3.9961>.
- Kelly, R. P., D. M. Lodge, K. N. Lee, et al. 2024. "Toward a National eDNA Strategy for the United States." *Environmental DNA* 6: e432. <https://doi.org/10.1002/edn3.432>.
- Kim, Y., and J. M. Stanton. 2016. "Institutional and Individual Factors Affecting Scientists' Data-Sharing Behaviors: A Multilevel Analysis." *Association for Information Science and Technology* 67: 776–799. <https://doi.org/10.1002/asi.23424>.
- Kimble, M., S. Allers, K. Campbell, et al. 2022. "Medna-Metadata: An Open-Source Data Management System for Tracking Environmental DNA Samples and Metadata." *Bioinformatics* 38: 4589–4597. <https://doi.org/10.1093/bioinformatics/btac556>.
- Klymus, K. E., J. D. Baker, C. L. Abbott, et al. 2024. "The MIEM Guidelines: Minimum Information for Reporting of Environmental Metabarcoding Data." *Metabarcoding and Metagenomics* 8: e128689.
- Klymus, K. E., C. M. Merkes, M. J. Allison, et al. 2020. "Reporting the Limits of Detection and Quantification for Environmental DNA Assays." *Environmental DNA* 2: 271–282. <https://doi.org/10.1002/edn3.29>.
- Kodama, Y., T. Ara, A. Fukuda, et al. 2024. "DDBJ Update in 2024: The DDBJ Group Cloud Service for Sharing Pre-Publication Data." *Nucleic Acids Research* 53, no. D1: D45–D48. <https://doi.org/10.1093/nar/gkae882>.
- Leigh, D. M., A. G. Vandergast, M. E. Hunter, et al. 2024. "Best Practices for Genetic and Genomic Data Archiving." *Nature Ecology & Evolution* 8: 1224–1232. <https://doi.org/10.1038/s41559-024-02423-7>.
- Lesperance, M. L., M. J. Allison, L. C. Bergman, M. D. Hocking, and C. C. Helbing. 2021. "A Statistical Model for Calibration and Computation of Detection and Quantification Limits for Low Copy Number Environmental DNA Samples." *Environmental DNA* 3: 970–981. <https://doi.org/10.1002/edn3.220>.
- Liggins, L., M. Hudson, and J. Anderson. 2021. "Creating Space for Indigenous Perspectives on Access and Benefit-Sharing: Encouraging Researcher Use of the Local Contexts Notices." *Molecular Ecology* 30: 2477–2482. <https://doi.org/10.1111/mec.15918>.
- Local Contexts. 2023. "Grounding Indigenous Rights." <https://localcontexts.org/>.
- Mandeville, C. P., W. Koch, E. B. Nilsen, and A. G. Finstad. 2021. "Open Data Practices Among Users of Primary Biodiversity Data." *Bioscience* 71: 1128–1147.
- Martone, M. E. 2015. "FORCE11: Building the Future for Research Communications and e-Scholarship." *Bioscience* 65: 635. <https://doi.org/10.1093/biosci/biv095>.
- Matthias, L., M. J. Allison, C. Y. Maslovat, J. Hobbs, and C. C. Helbing. 2021. "Improving Ecological Surveys for the Detection of Cryptic, Fossorial Snakes Using eDNA on and Under Artificial Cover Objects." *Ecological Indicators* 131: 108–187.
- McCartney, A. M., M. A. Head, K. S. Tsosie, et al. 2023. "Indigenous Peoples and Local Communities as Partners in the Sequencing of Global Eukaryotic Biodiversity." *NPJ Biodiversity* 2, no. 1: 8. <https://doi.org/10.1038/s44185-023-00013-7>.
- McKiernan, E. C., P. E. Bourne, C. T. Brown, et al. 2016. "How Open Science Helps Researchers Succeed." *eLife* 5: e16800. <https://doi.org/10.7554/eLife.16800>.
- McMurry, J. A., N. Juty, N. Blomberg, et al. 2017. "Identifiers for the 21st Century: How to Design, Provision, and Reuse Persistent Identifiers to Maximize Utility and Impact of Life Science Data." *PLoS Biology* 15: e2001414. <https://doi.org/10.1371/journal.pbio.2001414>.
- Meyer, R., W. Appeltans, W. Duncan, et al. 2023. "Aligning Standards Communities for Omics Biodiversity Data: Sustainable Darwin Core-MIx Interoperability." *Biodiversity Data Journal* 11: e112420. <https://doi.org/10.3897/BDJ.11.e112420>.
- Norris, E., and D. B. O'Connor. 2019. "Science as Behaviour: Using a Behaviour Change Approach to Increase Uptake of Open Science." *Psychology & Health* 34: 1397–1406. <https://doi.org/10.1080/08870446.2019.1679373>.
- NSF—National Science Foundation. 2024. "Proposal & Award Policies & Procedures Guide (PAPP) (NSF 24–1)." <https://new.nsf.gov/policies/pappg/24-1>.
- O'Brien, J. M., N. Blais, C. Butler, et al. 2024. "Ten "Simple" Rules for Non-Indigenous Researchers Engaging Indigenous Communities in Arctic Research." *PLoS Computational Biology* 20: e1012093.
- Open Research Europe. 2025. "Open Access Publishing Platform." <https://open-research-europe.ec.europa.eu/>.
- Parks, D. H., M. Chuvochina, C. Rinke, A. J. Mussig, P.-A. Chaumeil, and P. Hugenholtz. 2022. "GTDB: An Ongoing Census of Bacterial and Archaeal Diversity Through a Phylogenetically Consistent, Rank Normalized and Complete Genome-Based Taxonomy." *Nucleic Acids Research* 50: D785–D794. <https://doi.org/10.1093/nar/gkab776>.
- Piwowar, H. A., R. S. Day, and D. B. Fridsma. 2007. "Sharing Detailed Research Data Is Associated With Increased Citation Rate." *PLoS One* 2: e308. <https://doi.org/10.1371/journal.pone.0000308>.
- Piwowar, H. A., and T. J. Vision. 2013. "Data Reuse and the Open Data Citation Advantage." *PeerJ* 1: e175. <https://doi.org/10.7717/peerj.175>.
- R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.
- Ratnasingham, S., and P. D. N. Hebert. 2007. "Bold: The Barcode of Life Data System." *Molecular Ecology Notes* 7: 355–364. <https://doi.org/10.1111/j.1471-8286.2007.01678.x>.

- Reyes-García, V., R. Cámara-Leret, B. S. Halpern, et al. 2023. "Biocultural Vulnerability Exposes Threats of Culturally Important Species." *Proceedings of the National Academy of Sciences* 120: e2217303120. <https://doi.org/10.1073/pnas.2217303120>.
- Richardson, L., B. Allen, G. Baldi, et al. 2023. "MGnify: The Microbiome Sequence Data Analysis Resource in 2023." *Nucleic Acids Research* 51: D753–D759. <https://doi.org/10.1093/nar/gkac1080>.
- Roche, D. G., L. E. B. Kruuk, R. Lanfear, and S. A. Binning. 2015. "Public Data Archiving in Ecology and Evolution: How Well Are We Doing?" *PLoS Biology* 13: e1002295. <https://doi.org/10.1371/journal.pbio.1002295>.
- Rourke, M. L., A. M. Fowler, J. M. Hughes, et al. 2022. "Environmental DNA (eDNA) as a Tool for Assessing Fish Biomass: A Review of Approaches and Future Considerations for Resource Surveys." *Environmental DNA* 4: 9–33. <https://doi.org/10.1002/edn3.185>.
- Ruppert, K. M., R. J. Kline, and M. S. Rahman. 2019. "Past, Present, and Future Perspectives of Environmental DNA (eDNA) Metabarcoding: A Systematic Review in Methods, Monitoring, and Applications of Global eDNA." *Global Ecology and Conservation* 17: e00547.
- Saccò, M., M. T. Guzik, M. Van Der Heyde, et al. 2022. "eDNA in Subterranean Ecosystems: Applications, Technical Aspects, and Future Prospects." *Science of the Total Environment* 820: 153223. <https://doi.org/10.1016/j.scitotenv.2022.153223>.
- Samuel, R., R. Meyer, P. L. Buttigieg, et al. 2021. "Toward a Global Public Repository of Community Protocols to Encourage Best Practices in Biomolecular Ocean Observing and Research." *Frontiers in Marine Science* 8: 758694. <https://doi.org/10.3389/fmars.2021.758694>.
- Sayers, E. W., M. Cavanaugh, K. Clark, et al. 2024. "GenBank 2024 Update." *Nucleic Acids Research* 52: D134–D137. <https://doi.org/10.1093/nar/gkad903>.
- Shea, M. M., and A. B. Boehm. 2024. "Environmental DNA Metabarcoding Differentiates Between Micro-Habitats Within the Rocky Intertidal." *Environmental DNA* 6: e521. <https://doi.org/10.1002/edn3.521>.
- Shea, M. M., J. Kuppermann, M. P. Rogers, D. S. Smith, P. Edwards, and A. B. Boehm. 2023. "Systematic Review of Marine Environmental DNA Metabarcoding Studies: Toward Best Practices for Data Usability and Accessibility." *PeerJ* 11: e14993. <https://doi.org/10.7717/peerj.14993>.
- Sholler, D., K. Ram, C. Boettiger, and D. S. Katz. 2019. "Enforcing Public Data Archiving Policies in Academic Publishing: A Study of Ecology Journals." *Big Data & Society* 6, no. 2: 6258. <https://doi.org/10.1177/2053951719836258>.
- Soranno, P. A., and D. S. Schimel. 2014. "Macrosystems Ecology: Big Data, Big Ecology." *Frontiers in Ecology and the Environment* 12: 3. <https://doi.org/10.1890/1540-9295-12.1.3>.
- Stoeckle, M. Y., M. Das Mishu, and Z. Charlop-Powers. 2020. "Improved Environmental DNA Reference Library Detects Overlooked Marine Fishes in New Jersey, United States." *Frontiers in Marine Science* 7: 226. <https://doi.org/10.3389/fmars.2020.00226>.
- Takahashi, M., M. Saccò, J. H. Kestel, et al. 2023. "Aquatic Environmental DNA: A Review of the Macro-Organismal Biomonitoring Revolution." *Science of the Total Environment* 873: 162322. <https://doi.org/10.1016/j.scitotenv.2023.162322>.
- Tedersoo, L., R. Küngas, E. Oras, et al. 2021. "Data Sharing Practices and Data Availability Upon Request Differ Across Scientific Disciplines." *Scientific Data* 8: 192. <https://doi.org/10.1038/s41597-021-00981-0>.
- Tenopir, C., E. D. Dalton, S. Allard, et al. 2015. "Changes in Data Sharing and Data Reuse Practices and Perceptions Among Scientists Worldwide." *PLoS One* 10: e0134826. <https://doi.org/10.1371/journal.pone.0134826>.
- Teytelman, L., A. Stoliartchouk, L. Kindler, and B. L. Hurwitz. 2016. "Protocols.io: Virtual Communities for Protocol Development and Discussion." *PLoS Biology* 14: e1002538. <https://doi.org/10.1371/journal.pbio.1002538>.
- Thalinger, B., K. Deiner, L. R. Harper, et al. 2021. "A Validation Scale to Determine the Readiness of Environmental DNA Assays for Routine Species Monitoring." *Environmental DNA* 3: 823–836. <https://doi.org/10.1002/edn3.189>.
- Thompson, L. R., and P. Thielen. 2023. "Decoding Dissolved Information: Environmental DNA Sequencing at Global Scale to Monitor a Changing Ocean." *Current Opinion in Biotechnology* 81: 102936. <https://doi.org/10.1016/j.copbio.2023.102936>.
- Thomsen, P. F., J. Kielgast, L. L. Iversen, P. R. Møller, M. Rasmussen, and E. Willerslev. 2012. "Detection of a Diverse Marine Fish Fauna Using Environmental DNA From Seawater Samples." *PLoS One* 7, no. 8: e41732. <https://doi.org/10.1371/journal.pone.0041732>.
- Thomsen, P. F., and E. Willerslev. 2015. "Environmental DNA—An Emerging Tool in Conservation for Monitoring Past and Present Biodiversity." *Biological Conservation* 183: 4–18. <https://doi.org/10.1016/j.biocon.2014.11.019>.
- Thrall, P. H., J. Chase, J. Drake, et al. 2023. "From Raw Data to Publication: Introducing Data Editing at Ecology Letters." *Ecology Letters* 26: 829–830. <https://doi.org/10.1111/ele.14210>.
- Tsuji, S., Y. Iguchi, N. Shibata, I. Teramura, T. Kitagawa, and H. Yamanaka. 2018. "Real-Time Multiplex PCR for Simultaneous Detection of Multiple Species From Environmental DNA: An Application on Two Japanese Medaka Species." *Scientific Reports* 8: 9138. <https://doi.org/10.1038/s41598-018-27434-w>.
- Villacorta-Rath, C., T. Espinoza, B. Cockayne, J. Schaffer, and D. Burrows. 2022. "Environmental DNA Analysis Confirms Extant Populations of the Cryptic Irwin's Turtle Within Its Historical Range." *BMC Ecology and Evolution* 22: 57. <https://doi.org/10.1186/s12862-022-02009-6>.
- Vines, T. H., R. L. Andrew, D. G. Bock, et al. 2013. "Mandated Data Archiving Greatly Improves Access to Research Data." *FASEB Journal* 27: 1304–1308. <https://doi.org/10.1096/fj.12-218164>.
- Weigand, H., A. J. Beerhmann, F. Čiampor, et al. 2019. "DNA Barcode Reference Libraries for the Monitoring of Aquatic Biota in Europe: Gap-Analysis and Recommendations for Future Work." *Science of the Total Environment* 678: 499–524. <https://doi.org/10.1016/j.scitotenv.2019.04.247>.
- West, K. M., Z. T. Richards, E. S. Harvey, R. Susac, A. Greal, and M. Bunce. 2020. "Under the Karst: Detecting Hidden Subterranean Assemblages Using eDNA Metabarcoding in the Caves of Christmas Island, Australia." *Scientific Reports* 10: 21–479. <https://doi.org/10.1038/s41598-020-78525-6>.
- White, N. E., M. T. Guzik, A. D. Austin, et al. 2020. "Detection of the Rare Australian Endemic Blind Cave Eel (*Ophisternon candidum*) With Environmental DNA: Implications for Threatened Species Management in Subterranean Environments." *Hydrobiologia* 847: 3201–3211. <https://doi.org/10.1007/s10750-020-04304-z>.
- Whitlock, M. C. 2011. "Data Archiving in Ecology and Evolution: Best Practices." *Trends in Ecology & Evolution* 26: 61–65. <https://doi.org/10.1016/j.tree.2010.11.006>.
- Whitlock, M. C., M. A. McPeck, M. D. Rausher, L. Rieseberg, and A. J. Moore. 2010. "Data Archiving." *American Naturalist* 175: 145–146. <https://doi.org/10.1086/650340>.
- Wieczorek, J., D. Bloom, R. Guralnick, et al. 2012. "Darwin Core: An Evolving Community-Developed Biodiversity Data Standard." *PLoS One* 7: e29715. <https://doi.org/10.1371/journal.pone.0029715>.
- Wilcox, T. M., K. S. McKelvey, M. K. Young, et al. 2020. "Parallel, Targeted Analysis of Environmental Samples via High-Throughput Quantitative PCR." *Environmental DNA* 2: 544–553. <https://doi.org/10.1002/edn3.80>.

- Wilkinson, M. D., M. Dumontier, I. J. Aalbersberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3, no. 160: 018. <https://doi.org/10.1038/sdata.2016.18>.
- Wilkinson, S. P., A. A. Gault, S. A. Welsh, et al. 2024. "TICI: A Taxon-Independent Community Index for eDNA-Based Ecological Health Assessment." *PeerJ* 12: e16963. <https://doi.org/10.7717/peerj.16963>.
- Wood-Charlson, E. M., Z. Crockett, C. Erdmann, A. P. Arkin, and C. B. Robinson. 2022. "Ten Simple Rules for Getting and Giving Credit for Data." *PLoS Computational Biology* 18: e1010476. <https://doi.org/10.1371/journal.pcbi.1010476>.
- Yilmaz, P., R. Kottmann, D. Field, et al. 2011. "Minimum Information About a Marker Gene Sequence (MIMARKS) and Minimum Information About any (x) Sequence (MIxS) Specifications." *Nature Biotechnology* 29: 415–420. <https://doi.org/10.1038/nbt.1823>.
- Young, M. K., D. J. Isaak, M. K. Schwartz, et al. 2018. "Species Occurrence Data From the Aquatic eDNAAtlas Database."
- Yuan, D., A. Ahamed, J. Burgin, et al. 2024. "The European Nucleotide Archive in 2023." *Nucleic Acids Research* 52: D92–D97. <https://doi.org/10.1093/nar/gkad1067>.

### Supporting Information

Additional supporting information can be found online in the Supporting Information section.