FAIRe guidelines

Best practice for making eDNA data FAIR

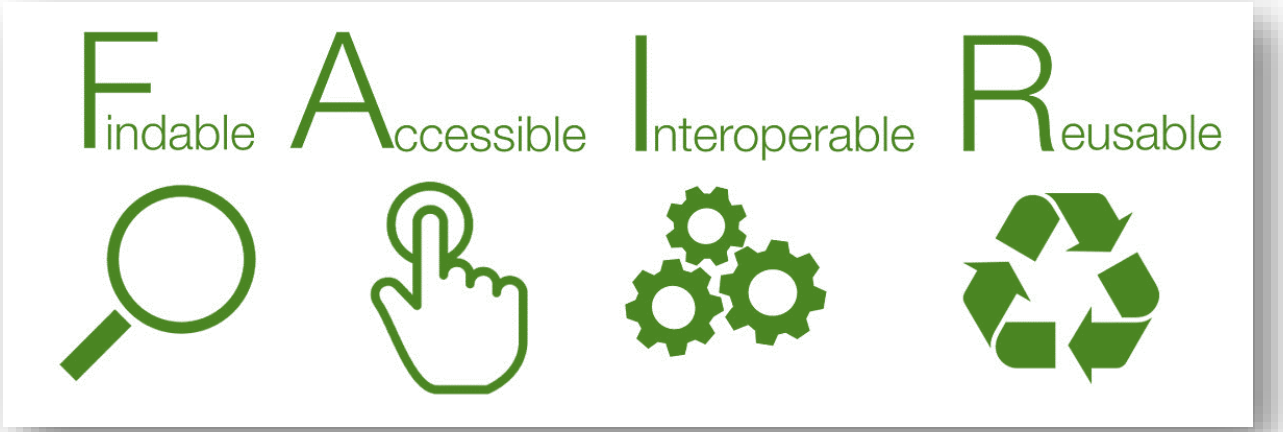
Miwa Takahashi & FAIRe working group

eDNA conference 2025 | Wellington, NZ

FAIRe website

Image generated by DALL-E

FAIR principles (Wilkinson et al., 2016)



Open data

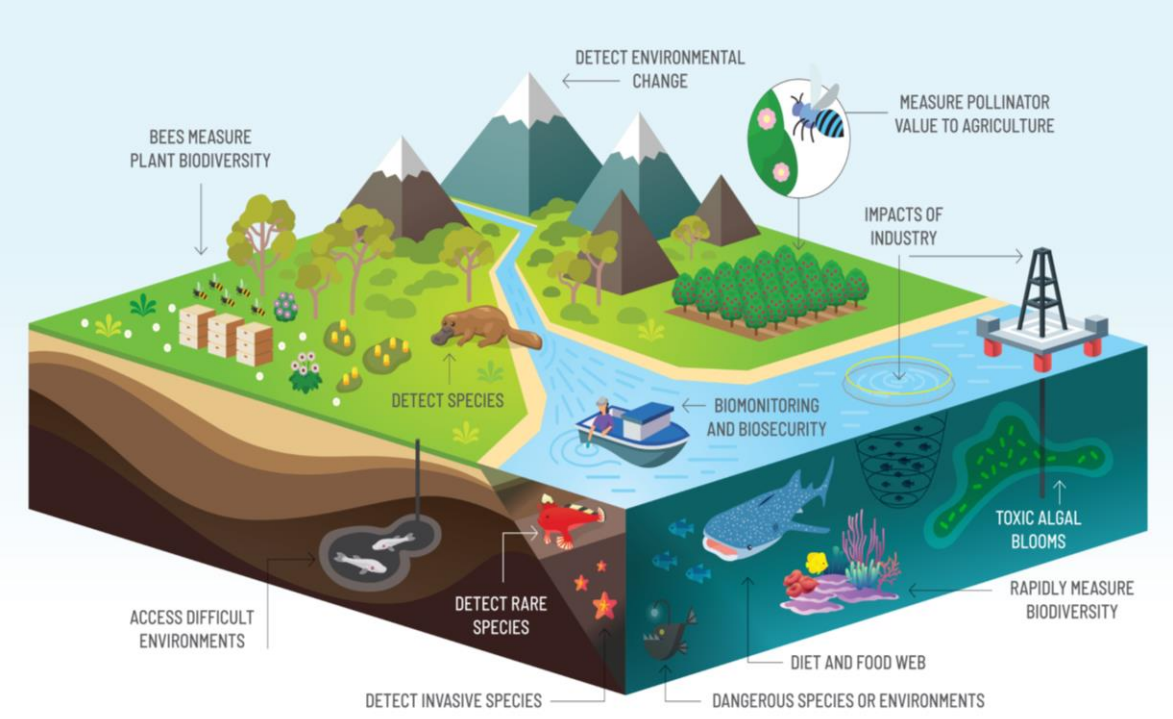
- Transparency
- Reproducibility

FAIR data

- Long-term value of data
- Generation of further knowledge at speed
- Cost-saving measure

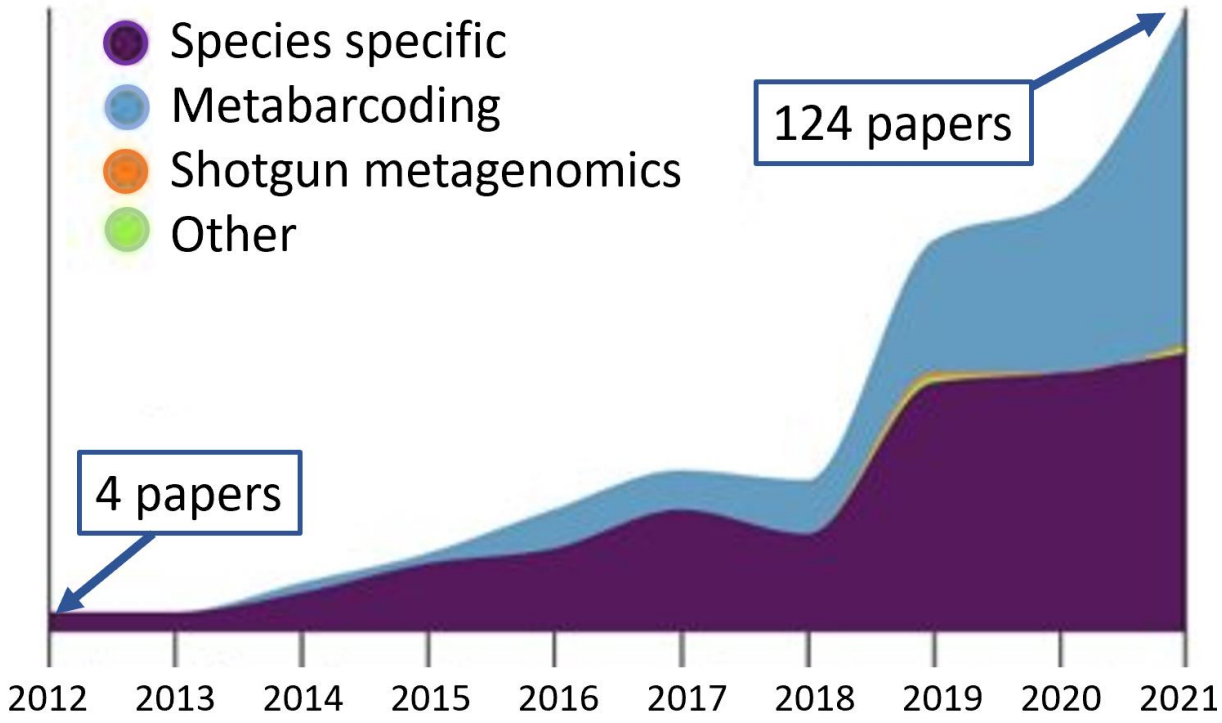
Making eDNA FAIR – Why? Why NOW?

eDNA applications



Berry et al. (2021)

# of aquatic eDNA studies on macrofauna



Takahashi & Sacco et al. (2023)

Benefits of FAIR eDNA

- Make studies more discoverable and cited

OPEN ACCESS Freely available online

PLOS one

Sharing Detailed Research Data Is Associated with Increased Citation Rate

Heather A. Piwowar*, Roger S. Day, Douglas B. Fridsma

Department of Biomedical Informatics, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, United States of America

PLOS ONE

RESEARCH ARTICLE

The citation advantage of linking publications to research data

Giovanni Colavizza[1,2], Iain Hrynaszkiewicz[3,4], Isla Staden[1,5], Kirstie Whitaker[1,6], Barbara McGillivray[1,6]*

1 The Alan Turing Institute, London, United Kingdom, 2 University of Amsterdam, Amsterdam, Netherlands, 3 Springer Nature, London, United Kingdom, 4 Public Library of Science, Cambridge, United Kingdom, 5 Queen Mary University, London, United Kingdom, 6 University of Cambridge, Cambridge, United Kingdom

PeerJ

Data reuse and the open data citation advantage

Heather A. Piwowar[1,2] and Todd J. Vision[1,2,3]

[1] National Evolutionary Synthesis Center, Durham, NC, USA
[2] Department of Biology, Duke University, Durham, NC, USA
[3] Department of Biology, University of North Carolina - Chapel Hill, Chapel Hill, NC, USA

Benefits of FAIR eDNA

- Make studies more discoverable and cited
- Run further studies at extended time/space
- Increase species ID accuracy
- Invasive species frontline monitoring
- Enhance collaboration

Generate knowledge at speed

Enhance data-driven decision making

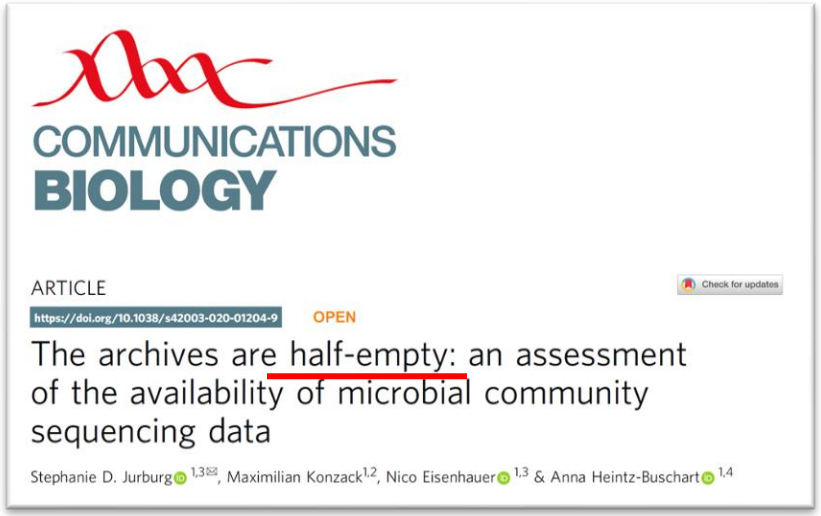
Current eDNA data: Open? FAIR?

Open, but not FAIR!! ☹

- Unstandardised format

  (e.g., Journal supplementary info, DRYAD , zenodo , etc)

- Incomplete data


COMMUNICATIONS BIOLOGY

ARTICLE

https://doi.org/10.1038/s42003-020-01204-9 OPEN

The archives are half-empty: an assessment of the availability of microbial community sequencing data

Stephanie D. Jurburg[1,3✉], Maximilian Konzack[1,2], Nico Eisenhauer[1,3] & Anna Heintz-Buschart[1,4]

FIAR eDNA (FAIRe) project

- FAIRe working group
  - 32 members, 25 institutions

CSIRO

2021 2030 United Nations Decade of Ocean Science for Sustainable Development

OBON OCEAN BIOMOLECULAR OBSERVING NETWORK

Biodiversity Information Standards TDWG

GBIF Global Biodiversity Information Facility

gsc genomic STANDARDS consortium

NGI Northern Gulf Institute

OBIS OCEAN BIODIVERSITY INFORMATION SYSTEM

Atlas of Living Australia ala.org.au

USGS science for a changing world

universität innsbruck

EMBL-EBI

ENA European Nucleotide Archive

UC SANTA BARBARA

NOAA National Oceanic and Atmospheric Administration U.S. Department of Commerce

TropWATER JAMES COOK UNIVERSITY AUSTRALIA

NATURAL ENGLAND

MUSÉUM NAL HIST NATURELLE

University of Victoria

WILDERLAB

Ministère des Forêts, de la Faune et des Parcs Québec

San Francisco State University

Environmental DNA Open Access
Dedicated to the study and use of environmental DNA for basic and applied sciences

WILEY

UNIVERSITY OF MARYLAND

FIAR eDNA (FAIRe) project

- FAIRe working group
  - 32 members, 25 institutions
- Review existing data standards
- Identify gaps
- Fill the gaps → FAIRe guidelines!

E.g.,

- Darwin Core (DwC) = Biodiversity data standard

Biodiversity Information Standards TDWG

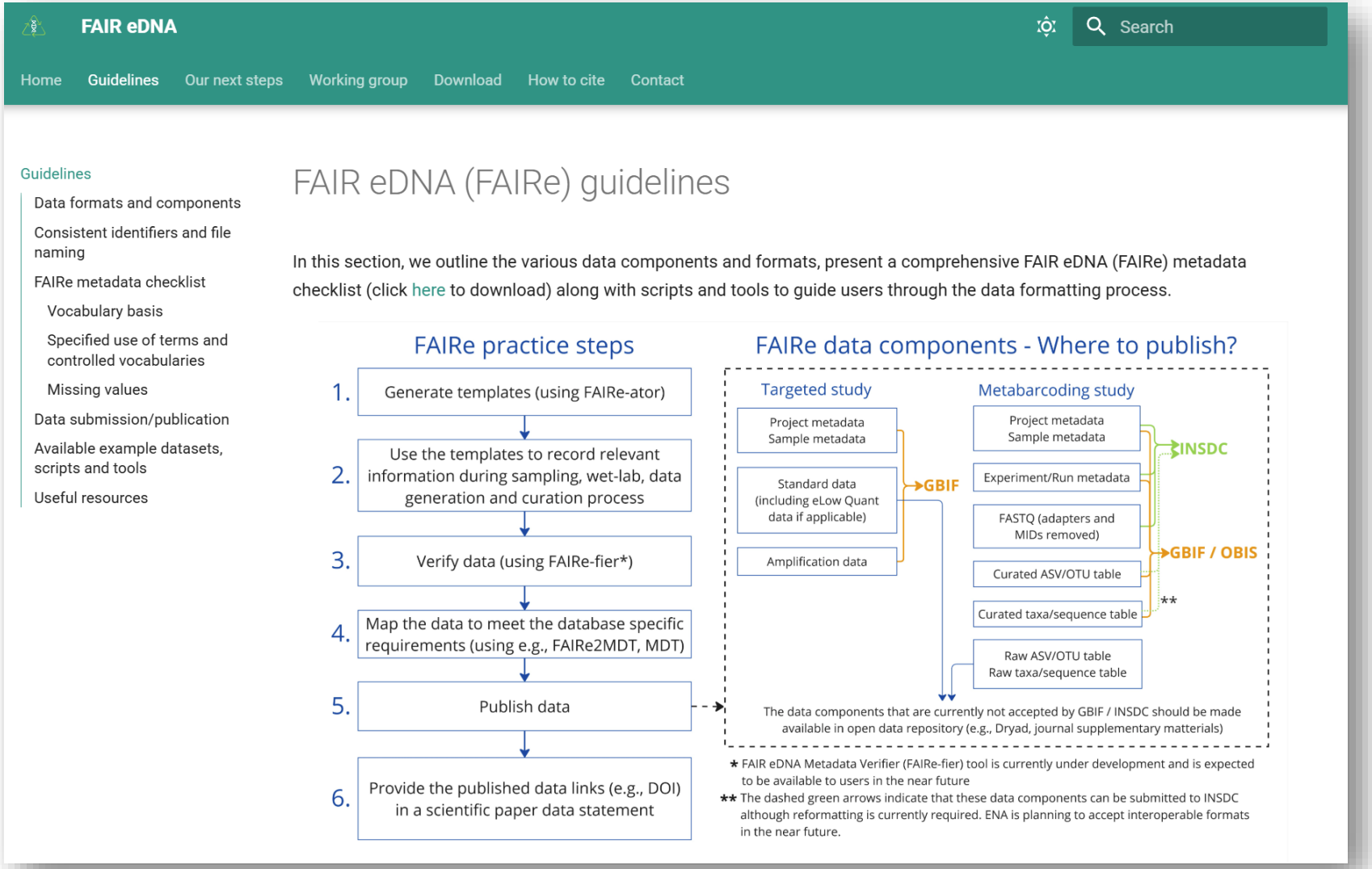
- Minimum Information about any (x) Sequences (MIxS) = Molecular data standard

gsc genomic STANDARDS consortium

- DNA-derived biodiversity data extension

GBIF

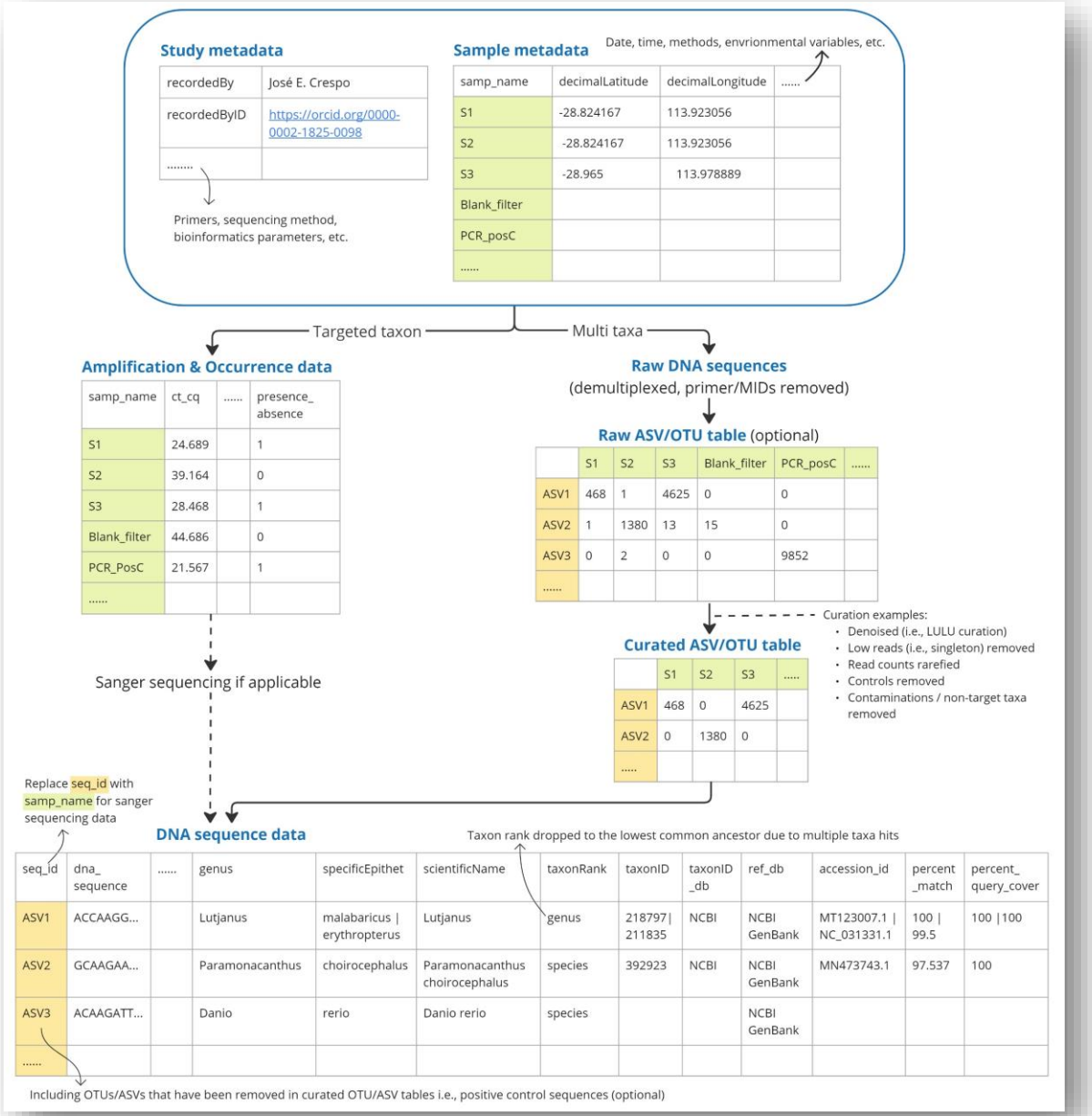
FAIRe guidelines



FAIRe website (https://fair-edna.github.io/index.html)

FAIRe guidelines

➢ Data components and formats



FAIRe guidelines

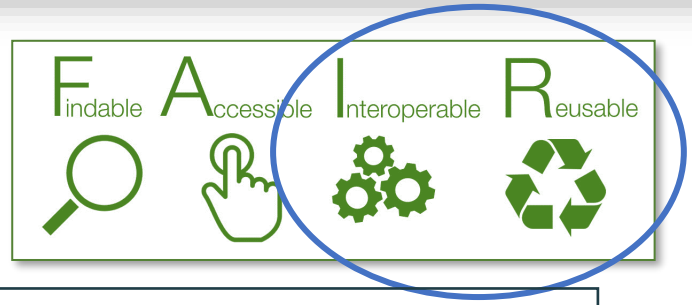
➢ Metadata checklist

section	term_name	description	requirement_level	term_type
Sample collection	**samp_collect_device**	The device used to collect an environmental sam	Recommended	free text
Sample collection	**samp_size**	Amount or size of original sample (volume, mass	Highly recommended	numeric
Sample collection	**samp_size_unit**	Unit for samp_size	Highly recommended	controlled vocabulary
Sample collection	**biological_rep**	The numbe of biological replicated collected at e	Recommended	integer
Sample storage	**samp_store_temp**	Temperature at which the sample was stored. En	Highly recommended	numeric or controlled vocabulary
Sample storage	**samp_store_sol**	Solution within which the sample was stored. Nc	Highly recommended	controlled vocabulary
Sample storage	**samp_store_dur**	Duration for which the sample was stored prior t	Highly recommended	fixed format
Sample storage	**samp_store_loc**	Location at which sample was stored, usually na	Optional	free text
Sample storage	**dna_store_loc**	Location at which extracted DNA was stored, usu	Optional	free text
Sample storage	**samp_store_method_a**	Additional information on sample storage metho	Optional	free text
Sample preparation	**samp_mat_process**	Any processing applied to the sample during or a	Recommended	free text
Sample preparation	**filter_passive_active_C**	Whether water/ air samples were collected using	Recommended	Boolean
Sample preparation	**stationed_sample_dur**	Duration of on-site, stationed sampling (i.e., the	Recommended	fixed format

Key focuses:

✓ Complete information

✓ Machine readable



- 337 terms (38 mandatory terms)
- Built from existing standards

  e.g., gsc genomic STANDARDS consortium Biodiversity Information Standards TDWG
- 158 new terms

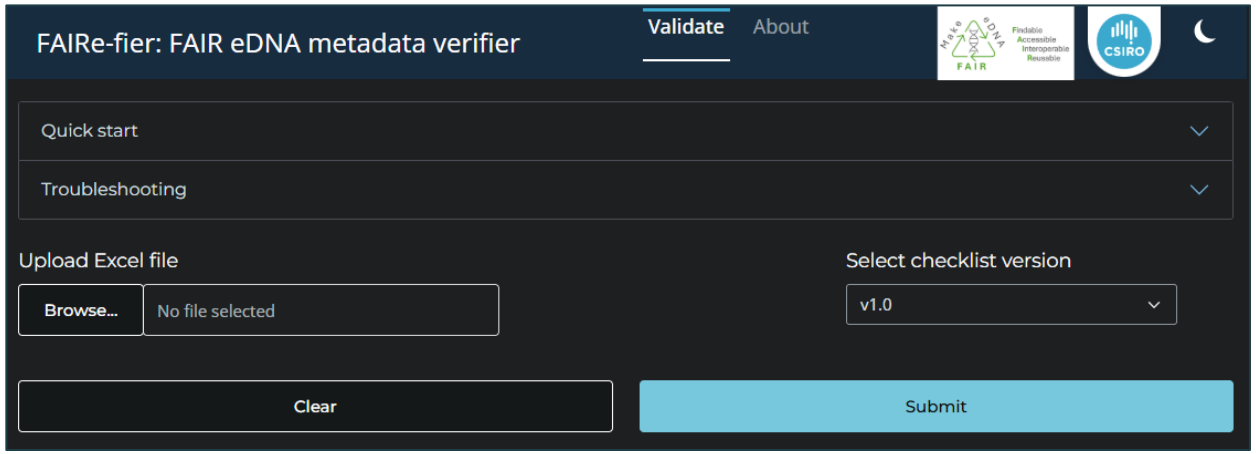
FAIRe guidelines

➢ Tools and scripts

- FAIRe-ator (FAIR eDNA template generator)

```
FAIReator(sample_type = c("Water", "Air"),
          assay_type = "metabarcoding",
          project_id = "gbr2022",
          assay_name = c("MiFish", "Crust16s"))
```

- FAIRe-fier (FAIR eDNA metadata verifer) – Coming soon!



- FAIRe2MDT (Metabarcoding data toolkit)

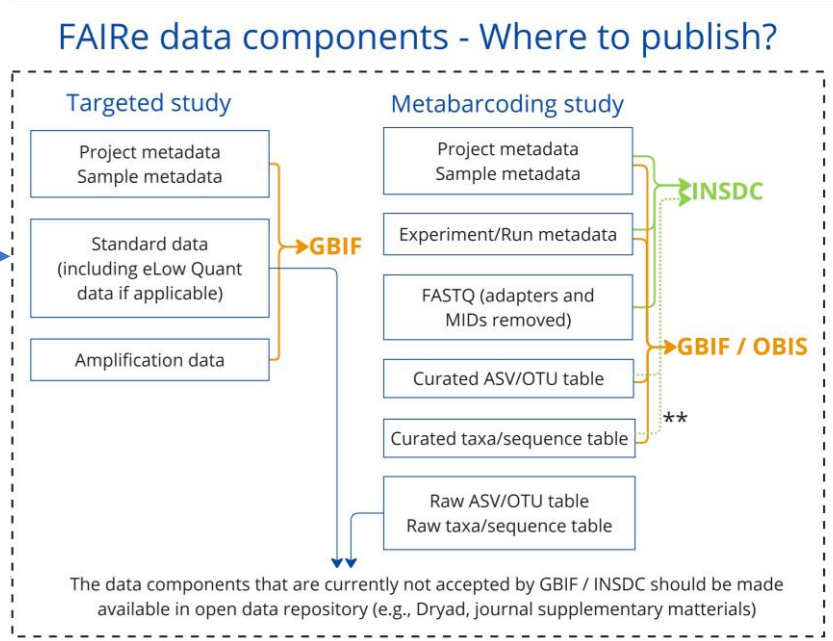
→ GBIF Global Biodiversity Information Facility

https://mdt.gbif.org/

FAIRe guidelines

➢ FAIRe practice steps 1 - 6

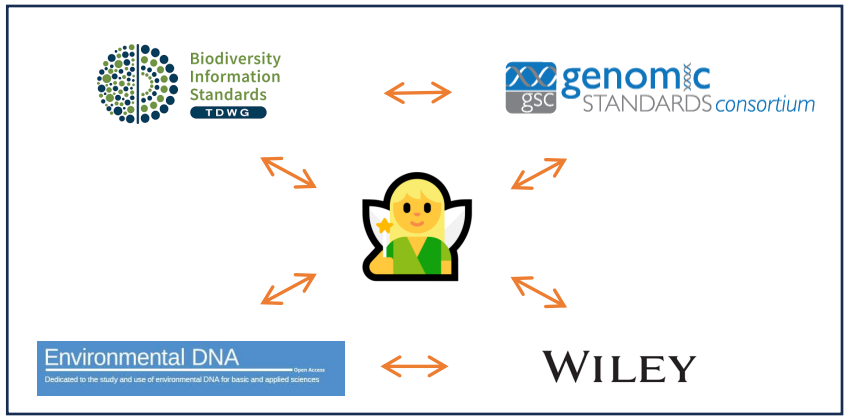
1. Generate templates (using FAIRe-ator)
2. Use the templates to record relevant information during sampling, wet-lab, data generation and curation process
3. Verify data (using FAIRe-fier*)
4. Map the data to meet the database specific requirements (using e.g., FAIRe2MDT, MDT)
5. Publish data
6. Provide the published data links (e.g., DOI) in a scientific paper data statement



What's next?

- Raising awareness
- Revising, updating the guidelines
- Integrating the guidelines
- Harmonising FAIR and CARE

1st FAIRe workshop at:



https://www.gida-global.org/care

Thank you!!

FAIRe Working Group

Miwa Takahashi
Olly Berry
Cecilia Villacorta Rath
Luke Thompson
Bruce Deagle
Erin Hahn
Katrina West
Shaun Wilkinson
Tobias Frøslev
Bettina Thalinger
Joana Pauperio
Katy Klymus
Gareth Jenkins
Martin Laporte
Christopher Jerde
Suk Yee Yong
Sachit Rajbhandari

Andrew Bissett
Peggy Newman
Thomas Stjernegaard Jeppesen
Katherine Silliman
Sean Jungbluth
Lynn Schriml
Chris Hunter
Steve Formel
Peter Woollard
Caren Helbing
Rachel Haderlé
Lynsey Harper
Nicholas Dunn
Neha Acharya-Patel
Mark Louie D. Lopez
Guy Cochrane