

Biodiversity data derived from DNA - How do we make it FAIR?

Oliver Williams | Senior Data Manager (Cefas) | oliver.williams@cefas.gov.uk

Kevin Paxman | Data Engineer (DASSH / MBA) | kevpax@mba.ac.uk

Dan Lear | Head Data, Information & Technology (DASSH / MBA) | dble@mba.ac.uk

Biodiversity occurrences derived from DNA are becoming an important source of data for many science areas. Within the Marine Environmental Data and Information Network (MEDIN) we are working towards a standardised approach for handling and sharing eDNA data biodiversity data products. How can the MEDIN community best reach out to our wider science colleagues to agree on shared principles, standards, and publication pipelines? We are keen to share our own ideas and learn from other networks. By working together we can ensure that DNA derived biodiversity data are open and FAIR; Findable, Accessible, Interoperable and Re-Usable across all networks and disciplines.

WHAT & WHO?

FAIR Data

Ensuring that data are [Findable, Accessible, Interoperable & Reusable \(FAIR\)](#)¹ helps improve re-use of science data and decreases mis-use. The FAIR data principles sit at the heart of many open science initiatives.

MEDIN

The [Marine Environmental Data & Information Network \(MEDIN\)](#) is a partnership of UK organisations committed to improving access to and stewardship of, marine data.

DASSH

As part of the Marine Biological Association, [DASSH](#) provides tools and services for the long-term curation, management, preservation and publication of marine species and habitats data, within the UK and internationally.

Cefas

As an executive agency of Defra, the [Centre for Environment, Fisheries, and Aquaculture Science](#) undertakes fundamental research and monitoring of the marine and freshwater environment and provides advice to policy makers nationally & internationally.

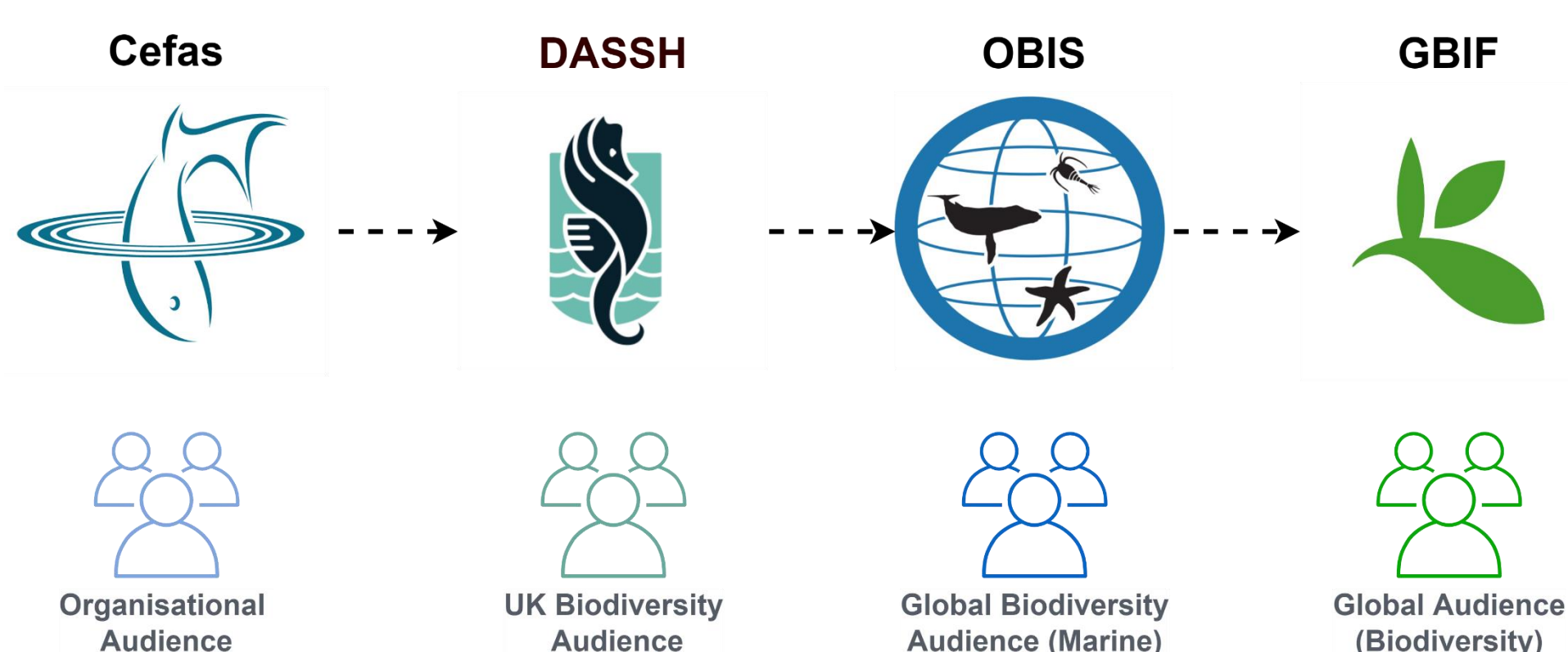
OBIS, GBIF & EMODnet

- [Ocean Biodiversity Information Facility \(OBIS\)](#) – a global data facility owned by the United Nations and focused on biodiversity in the Oceans.
- [Global Biodiversity Information Facility \(GBIF\)](#) – a cross governmental owned network & data infrastructure focused on global biodiversity.
- [European Marine Observation and Data Network \(EMODnet\)](#) – An EU funded network which provides access to data & data products. [EMODnet Biology](#) focuses on biodiversity and makes use of data flowing to OBIS & GBIF.

DATA FLOWS

- How do we share data without losing sight of it?
- How do we ensure scientists get credit for their work?
- How do users know who to contact with questions about the data?

The federated approach...



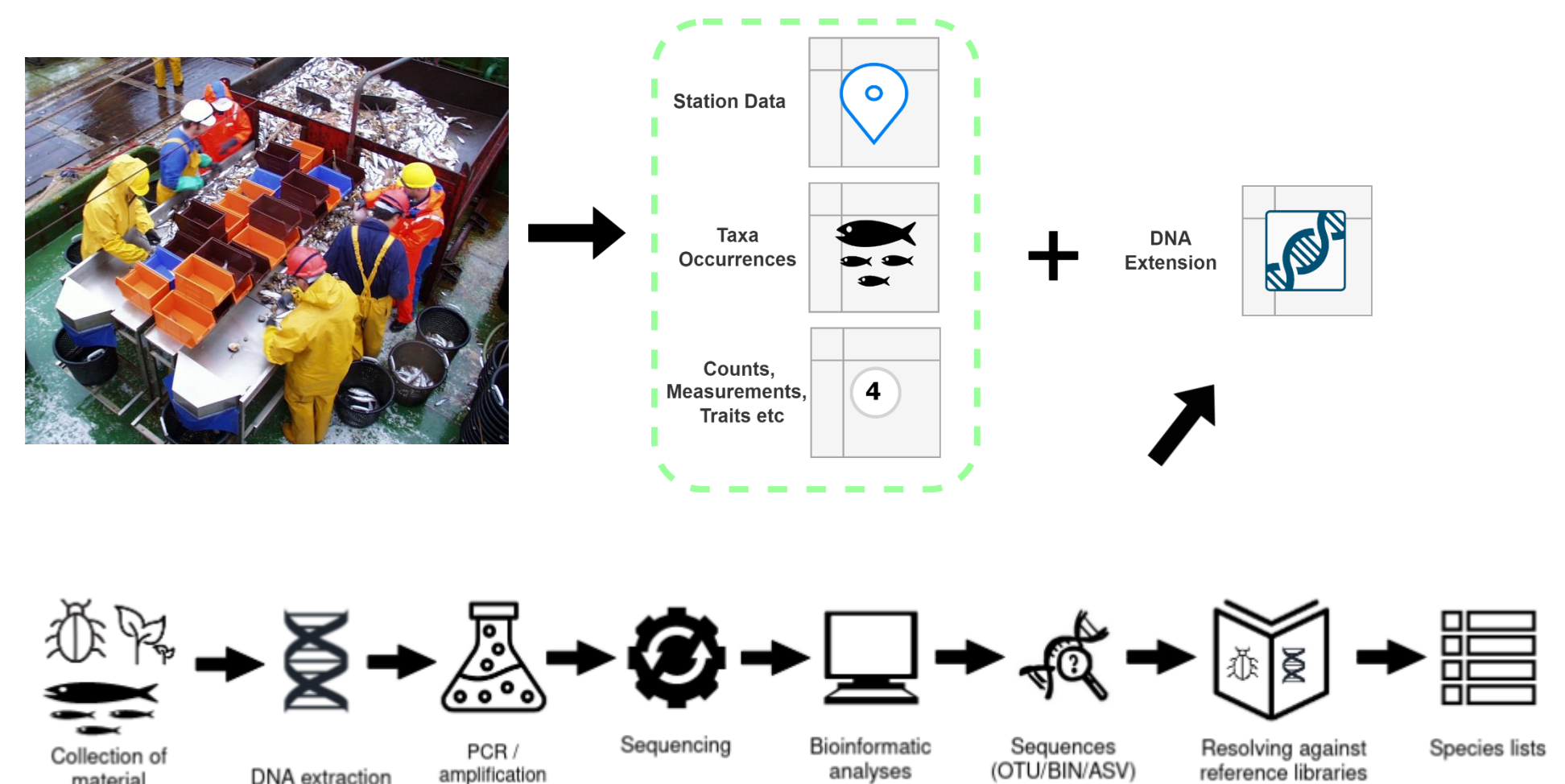
BIODIVERSITY OCCURRENCES

Occurrence data = Where? When? What?

In the seas, occurrence data for biodiversity may be collected by a large range different groups, for different purposes and often by different countries. It is a challenge to combine this data in an interoperable way.

The [OBIS](#) and [GBIF](#) use the [Darwin Core](#) format which allows data to be combined but without losing information on the exact way the taxa were observed.

An extension² has been created which incorporates DNA-derived data, might this work for you?



REACHING OUT

- Discussions continue within [MEDIN](#) and [EMODnet](#), how do we approach archival and sharing of Biodiversity Data derived from DNA?
- Are other networks and communities working on this?
- How do we better coordinate and learn from each other?
- Please reach out to us and our networks...



REFERENCES

1. Wilkinson *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>
2. 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>