DNA metadata reporting guidelines

For data not commissioned by Natural England

March 2025

Natural England Technical Information Note TIN224

Nick Dunn, Debbie Leatherland, Lynsey Harper

www.gov.uk/natural-england

NATURAL ENGLAND

Contents

Background

DNA data used by Natural England (NE) may enter the organisation in one of three ways:

1. NE commissions the collection and/or analysis of DNA data from an external contractor.

2. NE's partners commission the collection or analysis of DNA data from an external contractor, then provide this to NE to support monitoring or a specific project.

3. DNA data is submitted to NE as part of its function as a regulator, formally or informally.

This document is designed to provide guidance on the suggested metadata we recommend should be associated with DNA data for the data to be used for monitoring, casework and regulation by Natural England, where Natural England does not commission the collection or analysis **(categories 2 and 3 above)**. Where Natural England commissions the data (**category 1 above**), our standard metadata templates should be completed.

We will still use data provided to us in categories 2 and 3 without supporting metadata or with only some metadata, although our confidence in the data may be reduced and this may limit its use. To help ensure DNA-derived species data is of sufficient quality, and abides by FAIR (Findable, Accessible, Interoperable and Reusable) principles, sufficient metadata is required.

Metadata here is defined as; a set of data that describes and gives information about other data, it includes the necessary information needed to establish that the data has been produced using appropriate methods.

Box 1: Definition of metadata

Comprehensive metadata ensures that end users can understand the confidence they should have in the data. An added benefit is allowing the wider sharing and re-use of the data by the data owner. Natural England is currently creating guidance to help its staff who may receive DNA derived data understand how to best use the data, including understanding any accompanying metadata.

Where a company/organisation feels parts, or all, the metadata is considered commercial in confidence, please state that this is the case and provide as much detail as possible. We have identified areas which we believe may be commercial in confidence and have

provided suggestions for the information that may be possible to share. **This guidance is not mandatory, we are not mandating the sharing of information**, but trying to provide clarity on how data provided with limited metadata may be used. Here, we set out the metadata that we recommend should be provided alongside DNA-derived data.

Proposed metadata for single species DNA data

Reporting of the following information is recommended:

- **Sample date and location** including coordinates and the coordinate system used
- **Sample collection protocol**, including:
    - **Sample type**, e.g. water, soil, air, bulk DNA.
    - **DNA capture method**, e.g. filtration, precipitation.
    - **DNA capture medium**, e.g. filter type, material and pore size.

- **Sample size**, e.g. volume filtered, soak time or quantity of material sampled.

- **Sample preservation method** in field and in transit, including:
    - **Storage solution.**
    - **Storage temperature.**
    - **Storage time**.

- **Details of field and laboratory negative controls, and lab positive controls.**

- **DNA extraction method and protocol**, including (if commercial in confidence, please provide broad details on DNA extraction i.e. spin column or PCI method used):
    - **Any modifications to the protocol from a commercial kit or scientific publication.**
    - **Elution medium.**
    - **Elution volume.**
    - **Storage conditions of extracted DNA.**

- **Method used for determining concentration of extracted DNA, the DNA concentration of each sample and results of any quality testing on extracted DNA** e.g. Qubit, Nanodrop absorbance ratios

- **Details of any inhibition testing if carried out and attempts to reduce inhibition in DNA extracts**

- **Target gene and primer sequences (5'-3'), and amplicon length.** Include reference if using published primers and validation level from Thalinger and others

(2021). If sequences are commercial in confidence, please state target gene and amplicon length.

- **Probe sequence (5'-3'), the reporter dye and quencher used (if applicable). Include** reference if using published probe. If sequence is commercial in confidence, please state this, and confirm the use of a probe.

- **Number of technical replicates run per sample.**

- **PCR reagent details**, including:
    - **Total PCR reaction volume.**
    - **Volume of DNA template.**
    - **Concentration and volume of primers (and probe).**
    - **Master Mix used and volume added to reaction, or the exact volumes and concentrations of reagents added to a custom reaction mixture including enzyme used.**

- **PCR conditions**, including:
    - **Denaturation, annealing and extension temperatures and times.**
    - **Number of cycles.**

- **The limit of detection (LOD) and limit of quantification (LOQ) of the assay**, and the methods and level of replication used to establish these.

In addition, the assay must be supported with the following:

- ***In silico* analysis**, with details including:
    - **Database and software used.**
    - **Location of primers and probe on a sequence of target species' DNA**. A UK sequence for the target species should be used where possible, otherwise specify why a UK sequence has not been used and the country of origin of the sequence.
    - **Position of mismatches between the primers, probe, and closely related non-target species' sequences.** UK sequences for non-target species should be used where possible, otherwise specify why a UK sequence has not been used and the country of origin of the sequence.

- ***In vitro* analysis,** with details including:
    - **Tests on DNA extracted from tissue from multiple specimens of the target species, including at least one UK specimen.** If not possible, specify why UK specimens have not been used, and give the country of origin of the specimen. Results should show a clear amplification, with appropriate negative controls showing no amplification.
    - **Tests on DNA extracted from tissue from species closely related to the target species and any co-occurring species that are shown to be a potential co-amplification risk in the *in silico* analysis.** Multiple

specimens of non-target species should be tested, including at least one UK specimen of each non-target species where possible. If not possible, specify why UK specimens have not been used, and give the country of origin of the specimen.
    - Results for non-target species should be negative with appropriate negative controls showing no amplification and a positive control of the target species showing correct amplification.

- **Field testing,** with details including:
    - **Tests on an environmental sample from a location where, and season when, the target species is known to be present.** Results should be positive for the target species, with appropriate negative controls showing no amplification, and a positive control of the target species' DNA showing correct amplification.
    - **Tests on an environmental sample from a location where, and season when (if applicable), the target species is known to be absent.** Results should be negative for the target species, with appropriate negative controls showing no amplification, and positive control of the target species' DNA showing correct amplification

Suggested data format for single species DNA data

Our preferred data format (other formats are acceptable) for single species DNA data is:

- An Excel or CSV file in long format with one row per replicate containing:
    - Sample ID
    - Replicate ID
    - Ct value, calculated concentration / copy number
- An occurrence table with qPCR scores (number of positive replicates per sample)

Box 2: Recommended data format for single species DNA data

Proposed metadata for metabarcoding DNA data

Reporting of the following information is recommended:

- **Sample date and location** including coordinates and the coordinate system used
- **Sample collection protocol**, including:
    - **Sample type**, e.g. water, soil, air, bulk DNA.
    - **DNA capture method**, e.g. filtration, precipitation.
    - **DNA capture medium**, e.g. filter type, material and pore size.

- **Sample size**, e.g. volume filtered, soak time or quantity of material sampled.

- **Sample preservation method** in field and in transit, including:
    - **Storage solution.**
    - **Storage temperature.**
    - **Storage time**.

- **Details of field and laboratory negative controls, and lab positive controls.**

- **DNA extraction method and protocol**, including (if commercial in confidence, please provide broad details on DNA extraction i.e. spin column or PCI method used):

- **Any modifications to the protocol from a commercial kit or scientific publication.**
    - **Elution medium.**
    - **Elution volume.**
    - **Storage conditions of extracted DNA.**

- **Method used for determining DNA concentration of extracted DNA, the DNA concentration of each sample and results of any quality testing on extracted DNA** e.g. Qubit, Nanodrop absorbance ratios

- **Details of any inhibition testing if carried out and attempts to reduce inhibition in DNA extracts**

- **Target gene, gene region (if applicable, e.g. 16S V3), and fragment/amplicon length.**

- **State if one step PCR, two step PCR or tagged PCR strategy was used**

- **Forward and reverse primer sequences (5'-3'), and overhang sequences.** Include reference if using published primers. If sequence is commercial in confidence, please state amplicon length.

- **Blocking primer sequence** (5’-3’, if applicable). If subject to commercial in confidence, state that a blocking primer was used and its taxonomic target for blocking.

- **PCR reagent details**, including (if commercial in confidence, please state this and provide as much detail as possible):
    - **Total PCR reaction volume.**
    - **Volume of DNA template.**
    - **Concentration and volume of primers.**
    - **Master Mix used and volume added to reaction, or the exact volumes and concentrations of reagents added to a custom reaction mixture including enzyme used.**

- **PCR conditions,** including (if commercial in confidence, please state this and provide as much detail as possible):
    - **Denaturation, annealing and extension temperatures and times.**
    - **Number of cycles.**

- **Storage conditions of amplified samples.**

- **Number of technical replicates run per sample and pooling process.**

- **Method used for indexing.**

- **Sequencer make and model.**

- **Reagent kit used, percentage of PhiX control (if applicable), and library concentration used for sequencing.**

- **Target sequencing depth.**

- **Bioinformatics pipeline used**, including
    - **Details of software and version.**
    - **Thresholds and filters used.**

- **Numbers and percentages of reads discarded and assigned to taxonomy during bioinformatics.**

- **Details of reference database(s) used, the version and the access date; where not open access, state reference library and methods used for curation.**

- **Taxonomic assignment method used and thresholds for species assignments.**

- **Details of any tests run on mock communities and/or *in silico* testing**, including:
    - **Species tested against.**

    - **Any species that were not amplified or suggested to not amplify through *in silico* analysis.**

- **Details of quality control and assurance undertaken throughout processing**, including:
    - **Results showing the correct amplification of positive and negative controls throughout the process.**
    - **Methods used to account for any amplification in negative controls.**

Suggested data format for metabarcoding DNA data

Our preferred format (other formats are acceptable) is an Excel or CSV file in long format with one sheet per marker and one row per ASV/OTU containing:

- Sample ID
- ASV/OTU sequence
- Taxonomic assignments
- Taxonomic ID score (% identity/confidence score)
- Read count per sample

Box 3: Recommended data format for metabarcoding DNA data

Suggested templates for both DNA metadata and data are available on Natural England's Access to Evidence catalogue. The metadata templates have been adapted from the FAIR eDNA metadata templates (FAIR eDNA) to cover Natural England's requirements.

References

Takahashi, M., Frøslev, T.G., Paupério, J., Thalinger, B., Klymus, K., Helbing, C.C., Villacorta-Rath, C., Silliman, K., Thompson, L.R., Jungbluth, S.P., Yong, S.Y., Formel, S., Jenkins, G., Laporte, M., Deagle, B., Rajbhandari, S., Jeppesen, T.S., Bissett, A., Jerde, G., Hahn, E., Schriml, L.M., Hunter, C., Newman, P., Woollard, P., Harper, L.R., Dunn, N., West, K., Haderlé, R., Wilkinson, S., Acharya-Patel, N., Lopez, M.L.D., Cochrane, G., Berry, O. (in review) A metadata checklist and data formatting guidelines to make eDNA FAIR (Findable, Accessible, Interoperable and Reusable) https://fair-edna.github.io/index.html (accessed 22/01/2025)

Thalinger, B.,Deiner, K., Harper, L., Rees, H., Blackman, R., Sint, D., Traugott, M., Goldberg, C., Bruce, K.. 2021. A validation scale to determine the readiness of environmental DNA assays for routine species monitoring Environmental DNA https://doi.org/10.1002/edn3.189

About Natural England

Natural England is here to secure a healthy natural environment for people to enjoy, where wildlife is protected and England's traditional landscapes are safeguarded for future generations.

Further Information

This report can be downloaded from the Natural England Access to Evidence Catalogue. For information on Natural England publications or if you require an alternative format, please contact the Natural England Enquiry Service on 0300 060 3900 or email enquiries@naturalengland.org.uk.

Citation

Dunn, N; Leatherland, D; and Harper, L. 2025. DNA metadata reporting guidelines - For data not commissioned by Natural England. *Technical Information Note* TIN224. Natural England.

Copyright

Catalogue code: TIN224