



MICROBE
the Microbiome
Biobanking (RI)
Enabler

Let us pave the way
for microbiome
biobanking



MICROBE training sheet

Guidance BioSamples

Introduction and summary.

MICROBE has selected a representative set of samples from soil, seed and marine environments during the implementation of the project workplan. A focused approach has been adopted to ensure that all required developments and assessments can be thoroughly carried out and documented, resulting in a robust and reliable set of solutions. The project aims to deliver a comprehensive data infrastructure for microbiome biobanking that captures all linkages between biological samples, associated analytical data, and derived resources (e.g. microbial strains or consortia). This interconnected framework is essential to ensure data provenance, traceability, and reproducibility. MICROBE leverages the BioSamples platform, providing access to a global infrastructure for standardized sample metadata registration and retrieval. By using BioSamples, the project ensures seamless connectivity and traceability across all samples and datasets, adhering to FAIR (Findable, Accessible, Interoperable, Reusable) principles and guaranteeing the long-term sustainability of the data beyond the project's lifetime. This framework is designed to be reusable for other consortia and research infrastructures and exemplifies best practice implementation.

Overview and background

The BioSamples database (<https://www.ebi.ac.uk/biosamples/>) at the European Bioinformatics Institute (EMBL-EBI) is a core infrastructure resource that provides a centralised platform for the storage, curation, and dissemination of biological sample metadata. BioSamples holds sample descriptions across diverse datasets, enhancing their reusability and integrity in accordance with the FAIR principles. BioSamples serves as the central place to store the sample metadata for multiple repositories and archives at EMBL-EBI, including the European Nucleotide Archive (ENA), ArrayExpress, and the European Phenome-Gnome Archive. Registration of sample metadata at the EMBL-EBI BioSamples database results in persistent identifiers (accessions) which can be used as part of submissions to other omics archives, as the European Nucleotide Archive (ENA), ArrayExpress, the European Phenome-Genome Archive (EGA), PRIDE, and MetaboLights.

Reusing BioSamples accessions links the different data types through the metadata, providing context and enabling cross-functional analyses.

All MICROBE samples (environmental samples, isolates and assembled consortia) have been uniquely accessioned within the BioSamples database, and are foundational entities to which all other datasets, repositories, data files and analyses are connected. MICROBE samples registered in BioSamples (currently 4362 samples) can be found [here](#)*. Biosample records are structured to link to related BioSamples (e.g. isolates and assembled consortia) and associated datasets, thereby ensuring the findability and accessibility of both samples and data. This structure also facilitates access for external partners, supporting the validation of methods developed within the project. As additional data is analysed and published, future microbiome-derived datasets can be seamlessly linked to registered samples, leveraging a robust, flexible data infrastructure.

(*<https://www.ebi.ac.uk/biosamples/samples?filter=attr%3Aproject+name%3AMICROBE>)

Guidance on microbiome samples

SAMPLE REGISTRATION

Submission to BioSamples requires the use of sample metadata checklists. These include the Genome Standards Consortium Minimum Information about any (X) Sequence (MIxS) checklists. You can find the checklists [here](https://www.ebi.ac.uk/ena/browser/checklists). (<https://www.ebi.ac.uk/ena/browser/checklists>)

Mandatory fields for the chosen checklist: Depending on the type of sample, you will be required to validate your data against a different MIxS checklist. Currently, within MICROBE the following checklists are being utilised:

- ERC000020 for seed samples
- ERC000022 for soil samples
- ERC000024 for marine samples.

Additional metadata fields, specific to an assay or project, can be added to the templates as required. The MICROBE soil-specific templates are included [here](https://drive.google.com/drive/folders/1hTXqOShk-0uJ9SzMm2pRNqPnb3CpWUEZ). (<https://drive.google.com/drive/folders/1hTXqOShk-0uJ9SzMm2pRNqPnb3CpWUEZ>)

BioSamples allows sample registration via two routes, each of which is appropriate for a different set of users and submission types.

1) Interactive Submissions involve downloading template spreadsheets that can be completed offline and uploaded to BioSamples. This may be the easier option for non-programmatic submitters.

2) Programmatic Submissions involve preparing your submissions as JSON files before submitting them to BioSamples via the JSON API.

Please see the BioSamples documentation for more details on these submission routes and for further guidance:

<https://read-docs-biosamples.readthedocs.io/en/latest/>

External databases

ENA: sequence data is submitted to the [European Nucleotide Archive](#)

Mgnify: metagenomic analysis is performed by the [MGnify team](#)

BioStudies: functional data

All data above is linked by the samples.

Advantages for users and stakeholders*

The primary users of BioSamples include researchers, scientists, and research consortia. Numerous large-scale projects have used BioSamples as a central hub to connect multi-omics data and to develop data portals based on the structural framework provided by BioSamples.

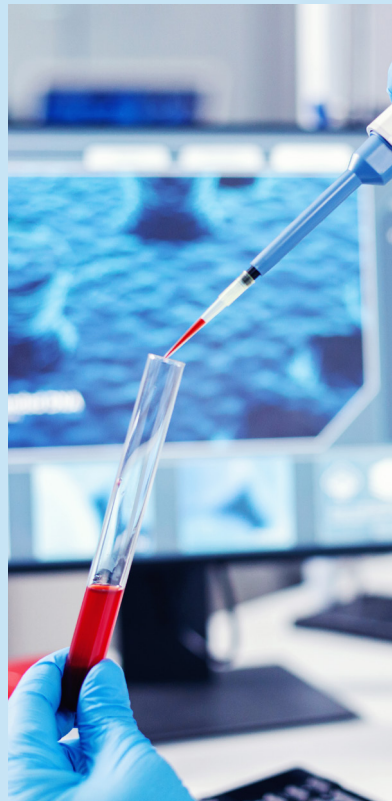
- Agriculture: [Holofoods](#), [FAANG](#)
- Marine: [Tara Oceans](#), [AtlantECO](#)
- Biodiversity: [Biodiversity portal](#) (includes projects like [Darwin Tree of Life](#), [European Reference Genome Atlas](#))
- Human Health: [COVID-19](#), [Pathogens](#), [Antimicrobial resistance](#), [HipSci](#), [MorPHiC](#)

Involved experts (partners)

EMBL, UK
MICROBE Consortium

Links & further information*

- [MICROBE samples in BioSamples](#)
- [MICROBE catalogue](#)
- [MICROBE submission Guide](#)
- [BioSamples documentation](#)
- [BioSamples publication](#)
- [BioSamples Webinar](#)



rtd services
innovation | cooperation | inspiration



HELMHOLTZ MUNICH



Contact

Biosamples Team
EMBL, UK
biosamples@ebi.ac.uk
www.ebi.ac.uk/biosamples

