

ENA – European Nucleotide Archive

[[About the Institution](#)] [[About the Data Center](#)] [[Data Center Profile](#)] [[Your contact persons at ENA](#)]

About the Institution

The [European Nucleotide Archive](#) provides a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation as well as metadata (sample description, experimental setup) and interpreted information (annotations). ENA is developed and operated by the [EMBL-European Bioinformatics Institute \(EMBL-EBI\)](#), an academic research institute based in the UK and part of the [European Molecular Biology Laboratory \(EMBL\)](#). ENA is one of the three databases that make up the [International Nucleotide Sequence Database Collaboration \(INSDC\)](#).



About the Data Center

The [European Nucleotide Archive](#) archives, curates and publishes nucleotide sequence data and associated metadata.

The GFBio Brokerage Service provides the timely, standards-compliant deposition of all molecular sequence data into the public repositories of the INSDC ([DDBJ](#), [EMBL-EBI](#) and [NCBI](#)). The key components of the service include: (a) Support for metadata standardization, curation and quality control, (b) negotiation of embargo periods, including communication with INSDC, (c) parallel submission of environmental metadata to [PANGAEA](#) and other [GFBio data centers](#), (d) cross-linking sequence and environmental data (e.g. [PANGAEA](#)) or other contextual or related data via accession number and DOI.



[Contact ENA at NFDI4Biodiversity](#)

Data Center Profile

Name	ENA – European Nucleotide Archive
URL	https://www.ebi.ac.uk/ena/browser/about
Description	<p>The EBI is a world-renowned center for research and services in bioinformatics and is the European node for globally coordinated efforts to collect and disseminate biological data. The EBI operates as the central hub of the intra-European infrastructure Elixir, whose goal is to orchestrate the collection, quality control and archiving of large amounts of biological data produced by life science experiments. As such, the EBI's mission is to ensure that the growing body of information from molecular biology and genome research is placed in the public domain and is accessible freely to all facets of the scientific community in ways that promote scientific progress. The ENA Team has almost 30 years of experience in capturing nucleotide sequence data, including the metadata that describes the experimental design for producing it. Currently ENA holds about 2.5 Petabyte of data with an approximate doubling time of 20 months.</p> <p>Scientific data curation services (incl. taxonomic services)</p> <p>Nucleotide sequence data and associated information (metadata) is deposited to the ENA using one of three submission routes: (1) programmatic submission, (2) interactive submission using the submission interface Webin, and (3) semi-automated route, where metadata are submitted using the interactive interface and data are deposited via an established institutional data submission service. Regardless of the submission route, all data and metadata are subject to the same validation tests. ENA issues permanent identifiers to all conceptual objects of the ENA data model and supports consistent description of the objects using checklists of information elements specific to each object.</p> <p>Scientific data curation is progressively moving from manual review of sequence annotation towards more impactful definition of validation rules for all supported data classes and checklists of the data objects. This approach allows more scalable and sustainable quality checks of incoming data. The ENA team also provides a helpdesk that supports depositors in resolving issues related to data submission as well as data retrieval.</p> <p>Taxonomic classification of all sequence data is based on the NCBI Taxonomy index and all incoming sequences are validated against this index. Organisms yet unclassified at the NCBI Taxonomy follow rules summarised at the ENA website. Essentially, data depositors report basic details on the unclassified sequenced organism and an amendment is requested at the NCBI Taxonomy, which is typically resolved within a few working days.</p> <p>ENA as a primary data archive serves data either for direct browsing, search or download or provides infrastructure services to domain-specific databases that add value to the primary data.</p>
Data domains (scope)	ENA's main focus is on long-term archiving and publication of nucleotide sequence data.
Target group	Individuals, scientists and researchers from national and international research groups and institutions aiming to archive and publish molecular data.

Service Description	The GFBio Brokerage Service provides the timely, standards-compliant deposition of all molecular sequence data into the public repositories of the INSDC. The key components of the service include: (a) Support for metadata standardization, curation and quality control, (b) negotiation of embargo periods, including communication with INSDC, (c) parallel submission of environmental metadata to PANGAEA, (d) cross-linking sequence data and environmental data (PANGAEA) via accession number and DOI.			
	IT services			
	ENA provides software tools for efficient sequence data submission and retrieval. For a complete overview, see http://www.ebi.ac.uk/ena/software .			
	<ul style="list-style-type: none">for data submission<ul style="list-style-type: none">Data submission websiteWeb Services for programmatic submissionWebin Data StreamerToolkit for sequence data compressionfor data retrieval<ul style="list-style-type: none">Advanced search websiteWeb Services for searching and retrieving dataWeb Services for sequence similarity searchesWeb Services for taxonomy lookup			
	User services			
	<ul style="list-style-type: none">User workshops on sequence data archiving and discoveryTraining MaterialsFAQHelpdesk			
Service Levels	Data Set x	Data Package x	Data management x	Research Objects
Data Formats				
Data Submission Formats	Data	Sequence data has to be in one of the formats supported by ENA .		
	Metadata	Molecular sequence metadata should be compliant with the standards of the "Minimum information about any (x) gene sequence" (MInxS). It can be put in manually, uploaded in a GCDJ/JSON format or as a tab-separated TSV. Appropriate templates are available for all formats.		
Data Accessibility	Public access points	ENA , data is exchanged daily with members of the INSDC Consortium , DNA Data Bank of Japan (DDBJ) and National Center for Biotechnology Information (NCBI)		
	Standardised exchange formats			
	Data formats	CSV, fastq, fasta, BAM, CRAM		
	Long-term availability	Unlimited		
Data Publication Services	Data Citation	ENA issues Accession Numbers , which are to be included as citations in publications, which use the respective datasets. Accession numbers are available for different granularity levels (e.g. study/datasets, samples, etc.). ENA recommends citing the study Accession Number (i.e. dataset identifier) throughout the text of the publication, more details under Citing ENA Data .		
	DOI	Accession Number (ACC)		
Archiving (RAW-data ingest, data, media)				
Licenses / Terms of Use	ENA does not currently have the possibility to provide a license in a structured and consistent way. EMBL-EBI, the institution providing ENA as a service, is setting out a roadmap to rationalise licence information on its data resources. For more information see EMBL-EBI licensing .			
Documentation	https://ena-docs.readthedocs.io/en/latest/index.html			
Computing center, external service provider				
name of the associated computing center (s), (commercial) service provider (s) and services provided				
Backup				
Computing center, external service provider				

Your contact persons at ENA

Data curator

- Jimena Linares, Ivaylo Kostadinov

Technical contact

- Ivaylo Kostadinov



Contact ENA at NFDI4Biodiversity

NFDI contact persons

- Ivaylo Kostadinov



Do you have questions, feedback or need help?

[Contact our Helpdesk](#) for direct support.

