EMBRC
EUROPEAN
MARINE
BIOLOGICAL
RESOURCE
CENTRE

EMBRC-ERIC Data Management Plan

April 2017

Revision	Date	Modification	Author
V1.0	April 2017	First draft	VLIZ

TABLE OF CONTENTS

1. FRAMEWORK AND SCOPE

A. Administrative information

Initiative	The European Marine Biological Resource Centre (EMBRC-ERIC)
Responsibles	EMBRC-ERIC Headquarters
Contact details	4 Place Jussieu, 75252 Paris Cedex 05 (FR) Tower 46/00, 1st floor, bureau 101 Letterbox 93 info@embrc.eu
Reference number/ID	EMBRC-DMP-V1.0-2017

B. Purpose of the Data Management Plan

EMBRC-ERIC acknowledges the importance of good data management of research data and the creation of a sound Data Management Plan (DMP) in anticipation of planned research activities. This DMP should be considered a living document that can be revised and will mature based on progressing insights into the nature of the data collections while taking into account developing standards and evolving data related initiatives.

Data often have a longer lifespan than the research project that creates them. Researchers may continue to work on data after funding has ceased, follow-up projects may analyze or add to the data, and data may be re-used by other researchers.

Well organized, well documented, preserved and shared data are invaluable to advance scientific inquiry and to increase opportunities for learning and innovation.

Good data management is fundamental for high quality research data and research excellence. Data management covers all aspects of handling, organizing, documenting and enhancing research data, and enabling their sustainability and sharing.

C. Research framework, activities and objectives of data collection

EMBRC-ERIC is a distributed research infrastructure (RI), integrating national Nodes with strong track records in research, training and servicing the science community (Figure 1).

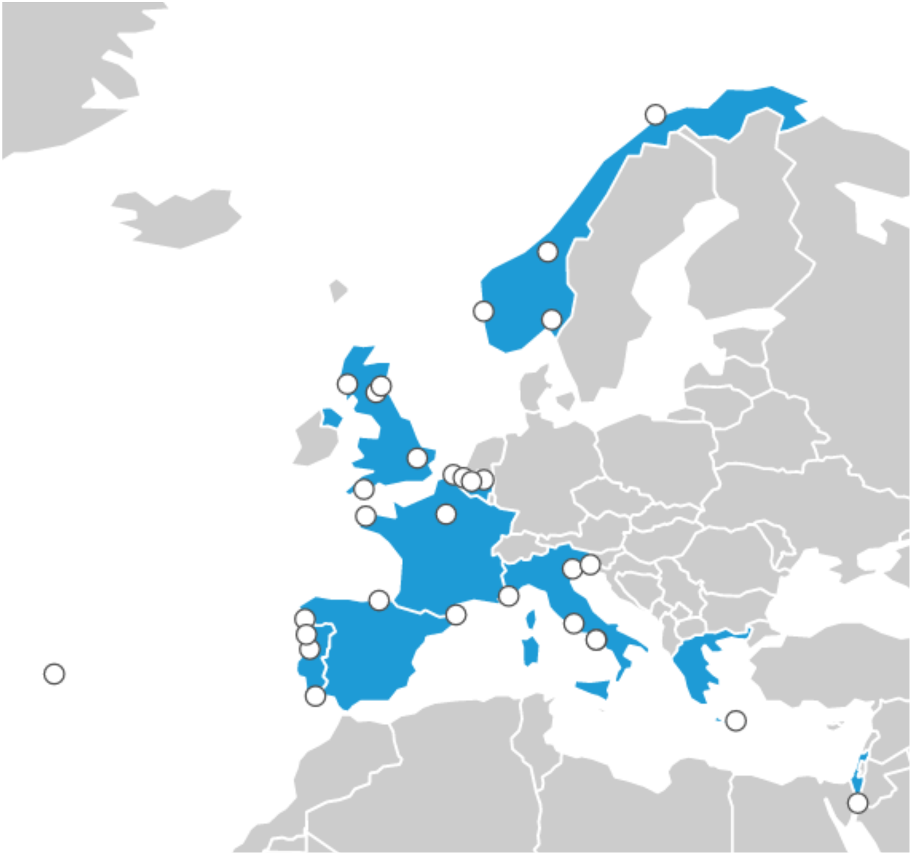

Figure 1 – Distribution of EMBRC-ERIC laboratories in Europe

These research laboratories host world-class in-house research communities and state-of-the-art infrastructure, with staff ranging from ca. 50-300 people. They share similar typologies, notably: (1) access to unique marine ecosystems and biological resources, including access to wet labs and culture collections, and (2) on-site support for genomics, post-genomics and bioinformatics. Many of the research communities have a high-impact track record of international collaboration involving several EMBRC-ERIC laboratories in various FP7, ERA and H2020 projects, having demonstrated excellence in science and strategic coordination over many years.

The EMBRC-ERIC infrastructure will enable marine biological research and industrial R&D by offering standardized access to a range of services:

- Marine ecosystems, including associated historical time-series data;
- Marine model organisms for academic and industrial research purposes;
- Logistics for ex-situ maintenance and experiments, including wet labs and up-to-date equipment for biological research ("omics");

- Rare and unique facilities for specialist research purposes, (e.g. bioreactors, micro- or mesocosms, marine mammal holding tanks, greenhouses);
- Biological and environmental data and bioinformatics;
- Teaching/training laboratory space and conference facilities, including logistics for hosting and catering visiting scientists.

The **EMBRC-ERIC activities** are depicted in Figure 2:

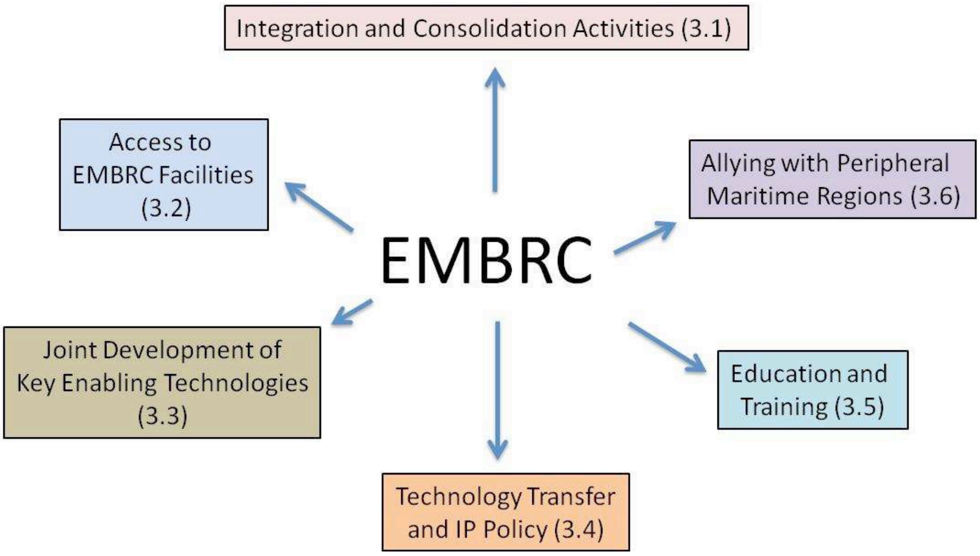
Integration and Consolidation Activities (3.1)
Access to EMBRC Facilities (3.2)
Allying with Peripheral Maritime Regions (3.6)
EMBRC
Joint Development of Key Enabling Technologies (3.3)
Education and Training (3.5)
Technology Transfer and IP Policy (3.4)

Figure 2 – Primary EMBRC-ERIC activities.

EMBRC-ERIC has a central position within the strategic research landscape, where services are extendable to biomedical sciences, environmental sciences and the biotechnology sector (Figure 3). These services cover both public and private sectors.

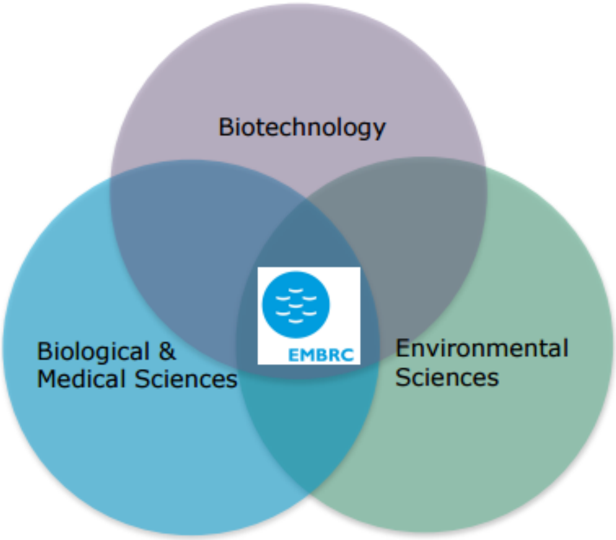

Figure 3 – EMBRC-ERIC central position within the strategic research landscape

D. Related documents, policies and procedures

Several documents contain statements that define the framework of this data management plan and therefore need to be referred to:

- The Statutes of the European Marine Biological Resource Centre – (EMBRC-ERIC Statutes)
- European Marine Biological Resource Centre (EMBRC) – Technical and Scientific Description (Final ERIC Application)
- EMBRC Scientific Strategy Report
- EMBRC-ERIC Data Policy (under development)
- EMBRC e-infrastructure strategy report (under development)

2. DATA COLLECTION

EMBRC-ERIC is a distributed Research Infrastructure that will provide users with access to its facilities and services. Through these activities EMBRC-ERIC will generate new research data. In addition EMBRC has the objective to provide virtual access to e-infrastructure and data. This chapter describes the EMBRC-ERIC research data collection generated by EMBRC operators and users. It explains the difference between background and foreground data, and gives an overview of the nature and types of data collected by EMBRC-ERIC, standards and methodologies in data creation, formats and software for sharing and long-term access to the data, file types for sharing, reuse and preservation of data; guidelines for structured data storage and quality assurance processes.

A. Background data versus foreground data

Based on the origin of the data, there are **3 general types of data** that can be considered to be part of the EMBRC data collection:

Type 1 data: data generated by a public research project carried out at EMBRC-ERIC originating from EMBRC user access provision and where the project covers the operating costs. These data are generated during on-site or remote access provision to a service or resource provided by the EMBRC operator, as part of the EMBRC service offer.

Type 2 data: data originating from projects that are undertaken under an "EMBRC umbrella". The origin of the projects' funding can vary (e.g. European, national, regional or local), but the projects and the role of the EMBRC partners involved are explicitly linked to EMBRC.

Type 3 data: EMBRC partners' institutional data that are considered to be part of the EMBRC service offer. These data are not necessarily generated in the context of EMBRC, but are explicitly listed by the EMBRC partners as part of the service offer.

Type 1 and 2 data can be considered as "foreground" data (i.e. data which is generated by EMBRC activities); Type 3 data can be considered as background data (e.g. long-term data that was gathered by the institute or station before the start of the EMBRC project).

B. Nature and types of data

Looking at the nature of the data, it is clear that each EMBRC operator is generating various types of data, resulting in an extremely diverse EMBRC research data collection. Table 1 gives an overview of the thematic data type categories generated within the different research domains of EMBRC-ERIC. For each data type category, the recommended data formats and data file types (see further) are listed.

	Thematic data type category	Research domains	Data formats	Data file types
Biological data	**Biodiversity data**	e.g. biogeographic research, species traits, taxonomy	Darwin Core Archive (DwC-A), OBIS-ENV-DATA	• Quanitative tabular data with minimal metadata:.csv; .tab; .xls; .xlsx, .txt; .mdb; .accdb; .dbf; .ods • Quantittatve tabular data with extensive metadata: .por; SPSS, Stata, Sas, DDI XML, .sav; .dta • Qualitative data: .xml; .rtf; .txt; .html; .doc; .docx; proprietary/software-specific formats, e.g. NUD*IST, NVivo and ATLAS.ti
	Genomic data	e.g. Sequencing (DNA, RNA,), annotation of features, protein structural information, gene expression profiles, alignment data, chromosomal mapping, phylogenetic trees, Single Nucleotide Polymorphisms (SNPs), functional genomics, Metabolomics, Proteomics, environmental DNA	• M2B3 Reporting Standard • Read data: general (CRAM, BAM, Fastq) and platform specific (SFF, PacBio, Oxford Nanopore, CompleteGenomics) • Assembled and annotated sequence data: flat file format (FASTA, XML), Multiple Sequence Alignment (MSA) formats	• Quanitative tabular data with minimal metadata:.csv; .tab; .xls; .xlsx, .txt; .mdb; .accdb; .dbf; .ods • Quantittatve tabular data with extensive metadata: .por; SPSS, Stata, Sas, DDI XML, .sav; .dta • Qualitative data: .xml; .rtf; .txt; .html; .doc; .docx; proprietary/software-specific formats, e.g. NUD*IST, NVivo and ATLAS.ti
	Imaging data	e.g. Zooscan images, flowcam images, flow cytometry images, VPR videos		• JPEG (.jpeg, .jpg), TIFF (.tif, .tiff), Adobe Portable Document Format (PDF/A, PDF) (.pdf), standard applicable RAW image format (.raw), Photoshop files (.psd), MPEG-4 (.mp4), motion JPEG 2000 (.mj2)
	Biogeochemical data	e.g. Biochemical pathways, Nutrients		• Quanitative tabular data with minimal metadata:.csv; .tab; .xls; .xlsx, .txt; .mdb; .accdb; .dbf; .ods • Quantittatve tabular data with extensive metadata: .por; SPSS, Stata, Sas, DDI XML, .sav; .dta • Qualitative data: .xml; .rtf; .txt; .html; .doc; .docx; proprietary/software-specific formats, e.g. NUD*IST, NVivo and ATLAS.ti
	Experimental data	e.g. Data resulting from lab experiments		• Quanitative tabular data with minimal metadata:.csv; .tab; .xls; .xlsx, .txt; .mdb; .accdb; .dbf; .ods

				• Quantittatve tabular data with extensive metadata: .por; SPSS, Stata, Sas, DDI XML, .sav; .dta • Qualitative data: .xml; .rtf; .txt; .html; .doc; .docx; proprietary/software-specific formats, e.g. NUD*IST, NVivo and ATLAS.ti
Oceanographic data	**Physical data**	e.g. Sea Water Temperature, Salinity, ocean currents, waves	NetCDF, Ocean Data View (ODV) format	• Quanitative tabular data with minimal metadata:.csv; .tab; .xls; .xlsx, .txt; .mdb; .accdb; .dbf; .ods • Quantittatve tabular data with extensive metadata: .por; SPSS, Stata, Sas, DDI XML, .sav; .dta • Qualitative data: .xml; .rtf; .txt; .html; .doc; .docx; proprietary/software-specific formats, e.g. NUD*IST, NVivo and ATLAS.ti
	Chemical data	e.g. Pollution, Heavy metals	NetCDF, Ocean Data View (ODV) format	• Quanitative tabular data with minimal metadata:.csv; .tab; .xls; .xlsx, .txt; .mdb; .accdb; .dbf; .ods • Quantittatve tabular data with extensive metadata: .por; SPSS, Stata, Sas, DDI XML, .sav; .dta • Qualitative data: .xml; .rtf; .txt; .html; .doc; .docx; proprietary/software-specific formats, e.g. NUD*IST, NVivo and ATLAS.ti
Climatological data	**Meteorological data**	e.g. Air temperature, Wind speed		• Quanitative tabular data with minimal metadata:.csv; .tab; .xls; .xlsx, .txt; .mdb; .accdb; .dbf; .ods • Quantittatve tabular data with extensive metadata: .por; SPSS, Stata, Sas, DDI XML, .sav; .dta • Qualitative data: .xml; .rtf; .txt; .html; .doc; .docx; proprietary/software-specific formats, e.g. NUD*IST, NVivo and ATLAS.ti
	Modelling data	e.g. Climate models		• Geospatial data (vector and raster data): .shp; .shx; .dbf; .prj; .sbx; .sbn; .tif; .tfw; .dwg; tabular GIS attribute data; .mdb; .mif; .kml; .ai; .dxf; .svg; binary formats of GIS and CAD packages

Table 1 – Overview of the thematic data type categories generated within the different research domains of EMBRC-ERIC. For each data type category, the recommended data file types, storage and redistribution channels (see further) are listed.

Furthermore, it should be noted that a distinction can be made between **raw data** and **processed and/or curated data**. Raw data, also known as primary data or unprocessed data, is a collection of numbers or characters before it has been cleaned, corrected or analyzed by researchers. Raw data needs to be corrected to remove outliers or obvious instrument or data entry errors. Data processing commonly occurs by stages, and the "processed data" from one stage may be considered the "raw data" of the next stage. Field data is raw data that is collected in an uncontrolled in-situ environment. Experimental data is data that is generated within the context of a scientific investigation by observation and recording (Source: Wikipedia).

C. Standards or methodologies in data creation

EMBRC acknowledges the fact that there are as many standards and methodologies in data creation as there are instruments and data types. Harmonization of procedures and standards in research activities and data collection is high on the priority list for EMBRC. At the moment, it is impossible to present an exhaustive list of methodologies applied in the various domains by the different EMBRC operators.

One of the infrastructure priorities identified in the EMBRC Scientific Strategy Report was the creation of a best practice database to achieve high and compatible standards in experimental methods, culture and husbandry, data collection and analysis This work has already been initiated as part of the Assemble project as a "virtual tool-box of best practice guide-lines" (http://www.assemblemarine.org/virtual-tool-box-of-best-practice-guide-lines/) (Annex 2). This toolbox aggregates a number of standard operating procedures, protocols, and guidelines on research activities, like holding, breeding and culturing of marine organisms, working with genetic and genomic resources, etc. During the follow-up project AssemblePlus, this toolbox will be updated with additional procedures and included as part of the EMBRC-ERIC Knowledge Transfer Platform. Specific joint research activity work packages deal with the harmonization of methodologies in relevant domains like: genomic observations, cryopreservation, functional genomics and experimental marine biology and ecology.

D. Formats and software for sharing and long-term access to the data

EMBRC-ERIC realizes that common data and metadata standards and formats are a key aspect for technological and semantic data operability in order to make data discoverable for promoting international and interdisciplinary access to and use of research data.

The format and software in which research data are created usually depend on how researchers choose to collect and analyze data, often determined by discipline-specific standards and customs. Ensuring long-term usability of data requires consideration of the most appropriate software and file formats. Given the broad spectrum of data types, there is a large offer of domain-specific formats for data sharing and often related thematic systems that can serve redistribution of those data. An important challenge for EMBRC will be to assess the different standards and workflows that will be proposed by the 3 cluster projects that it is involved in (EMBRIC, CORBEL, ENVRI+), and decide on common ones, both for its member institutes, but also with other RIs across domains to facilitate collaboration. In addition, the EMBRC Scientific Strategy report refers to the need for close collaboration with ESFRI initiatives, such as ELIXIR and LifeWatch, that will play a crucial role in determining how data is managed, shared and processed within EMBRC. These interactions have already been initiated by the EMBRC Working Group on E-Infrastructures, and EMBRC is preparing to increase this interaction through future activities.

Standardization at the data level will be performed applying community-based standards as proposed by international initiatives such as the International Oceanographic Data Exchange programme (IODE), the Ocean Biogeographic Information System (OBIS), the Genomics Standards Consortium (GSC) and the Open Geospatial Consortium (OGC). Notwithstanding the multitude of standards and formats available, EMBRC-ERIC already lists some of the recommended formats in the targeted domains.

Darwin Core Archive (DwC-A): The Darwin Core is designed to facilitate the exchange of information about the geographic occurrence of organisms and the physical existence of biotic specimens in collections. Extensions to the Darwin Core provide a mechanism to share additional information, which may be discipline-specific, or beyond the commonly agreed upon scope of the Darwin Core itself. The Darwin Core and its extensions are minimally restrictive of information content by design, since doing so would render the standard useless for the implementation of data quality tools (http://www.tdwg.org/activities/darwincore/).

OBIS-ENV-DATA: The OBIS-ENV-DATA format is a new data standard for combined marine biological and environmental datasets. It is based on Darwin Core Archive standard and recommended by the Ocean Biogeographic Information System part of the International Oceanographic Data Exchange network.
(http://bdj.pensoft.net/articles.php?id=10989).

M2B3 Reporting Standard: the M2BR standard was developed during the MB3 project to ensure that the collected data could be correctly directed to and stored in their respective domain-specific data archives, which were the ENA for molecular data and PANGAEA for environmental data and morphology-based biodiversity data.
(hdl.handle.net/10.1186/s40793-015-0001-5)

Data formats for read data (genomic data): Read data can be submitted in several standard (CRAM, BAM, Fastq) and platform specific formats (SFF, PacBio, Oxford Nanopore, Complete Genomics). ENA recommends that read data is either submitted in BAM or CRAM format (http://www.ebi.ac.uk/ena/submit/read-file-formats).

Data formats for assembled and annotated sequences (genomic data): The main format for assembled and annotated sequences is the flat file format, which is defined in full detail in the ENA Assembled Sequence User Manual. Assembled and annotated sequences are available in flat file and other formats, namely FASTA and XML, through the ENA Browser (http://www.ebi.ac.uk/ena/submit/sequence-format). Furthermore, Multiple Sequence Alignment (MSA) formats and General Feature Formats (GGF2, GGF3) for annotation exist.

NetCDF (oceanographic data): NetCDF (network Common Data Form) is a set of interfaces for array-oriented data access and a freely distributed collection of data access libraries for C, Fortran, C++, Java, and other languages. The NetCDF libraries support a machine-independent format for representing scientific data. Together, the interfaces, libraries, and format support the creation, access, and sharing of scientific data. (https://www.unidata.ucar.edu/software/netcdf/docs/faq.html#whatisit)

Ocean Data View (ODV) format (oceanographic data): The ODV data format allows dense storage and very fast data access. Large data collections with millions of stations can easily be maintained and explored on inexpensive desktop and notebook computers. Data from ARGO, GTSPP, CCHDO, World Ocean Database, World Ocean Atlas, World Ocean Circulation Experiment (WOCE), SeaDataNet, and Medar/Medatlas can be directly imported into ODV (https://odv.awi.de/). The British Oceanographic Data Centre distributes a SeaDataNet version of the general ODV format to carry additional information required by SeaDataNet (https://www.bodc.ac.uk/resources/delivery_formats/odv_format/).

See also Table 1 for the recommended data formats for the thematic data type categories within EMBRC-ERIC.

E. File types for sharing, reuse and preservation of data

Several data types require different file types. Table 2 gives an overview of possible file types for sharing, reuse and preservation (based on the UK Data Archive). EMBRC recommends to avoid the use of propriety software formats, since these are all temporary formats. If data files are in a different format, it is recommended to convert the data to a more common data format. See also Table 1 for the recommended data file types for the thematic data type categories within EMBRC-ERIC.

Table 2 - Overview of possible file types for sharing, reuse and preservation (UK Data Archive).

Type of data	Acceptable formats for sharing, reuse and preservation	Other acceptable formats for data preservation
Quantitative tabular data with extensive metadata a dataset with variable labels, code labels, and defined missing values, in addition to the matrix of data	• SPSS portable format (.por) • Delimited text and command ('setup') file (SPSS, Stata, SAS, etc.) containing metadata information • Some structured text or mark-up file containing metadata information, e.g. DDI XML file	• Proprietary formats of statistical packages e.g. SPSS (.sav), Stata (.dta) • MS Access (.mdb/.accdb)

Quantitative tabular data with minimal metadata a matrix of data with or without column headings or variable names, but no other metadata or labelling	• Comma-separated values (CSV) file (.csv) • Tab-delimited file (.tab) • Including delimited text of given character set with SQL data definition statements where appropriate	• Delimited text of given character set - only characters not present in the data should be used as delimiters (.txt) • Widely-used formats, e.g. MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb), dBase (.dbf) and OpenDocument Spreadsheet (.ods)
Geospatial data vector and raster data	• ESRI Shapefile (essential - .shp, .shx, .dbf, optional - .prj, .sbx, .sbn) • geo-referenced TIFF (.tif, .tfw) • CAD data (.dwg) • tabular GIS attribute data	• ESRI Geodatabase format (.mdb) • MapInfo Interchange Format (.mif) for vector data • Keyhole Mark-up Language (KML) (.kml) • Adobe Illustrator (.ai), CAD data (.dxf or .svg) • binary formats of GIS and CAD packages
Qualitative data textual	• eXtensible Mark-up Language (XML) text according to an appropriate Document Type Definition (DTD) or schema (.xml) • Rich Text Format (.rtf) • plain text data, ASCII (.txt)	• Hypertext Mark-up Language (HTML) (.html) • widely-used proprietary formats, e.g. MS Word (.doc/.docx) • some proprietary/software-specific formats, e.g. NUD*IST, NVivo and ATLAS.ti
Digital image data	• TIFF version 6 uncompressed (.tif)	• JPEG (.jpeg, .jpg) but only if created in this format • TIFF (other versions) (.tif, .tiff) • Adobe Portable Document Format (PDF/A, PDF) (.pdf) • standard applicable RAW image format (.raw) • Photoshop files (.psd)
Digital audio data	• Free Lossless Audio Codec (FLAC) (.flac)	• MPEG-1 Audio Layer 3 (.mp3) but only if created in this format • Audio Interchange File Format (AIFF) (.aif) • Waveform Audio Format (WAV) (.wav)
Digital video data	• MPEG-4 (.mp4) • motion JPEG 2000 (.mj2)	
Documentation and scripts	• Rich Text Format (.rtf) PDF/A or PDF (.pdf) HTML (.htm) OpenDocument Text (.odt)	• plain text (.txt) • some widely-used proprietary formats, e.g. MS Word (.doc/.docx) or MS Excel (.xls/.xlsx) • XML marked-up text (.xml) according to an appropriate DTD or schema, e.g. XHMTL 1.0

F. Structured data storage

EMBRC-ERIC will advise and train its members to organize a structured data storage. A structured data storage is essential for proper and secure storage of data files and records. For any file-based storage this includes clear and unambiguous file naming, the use of proper versioning, clear and intuitive folder structure. Data records that originate from different files but require integration can best be stored using relational databases.

Some recommendations for good file names:

- Create meaningful names (file names can contain e.g. project acronyms, researchers' initials, file type information, a version number, file status information and date);
- Use file names to classify broad types of files;
- Avoid using spaces and special characters;
- Avoid very long file names;

(From the UK Data Archive)

Version numbering in file names is useful to indicate files revisions or edits, especially in collaborations. This can be through discrete or continuous numbering depending on minor or major revisions.

It is important to think carefully how best to **structure files** in folders, in order to make it easy to locate and organize files and versions. When working in collaboration the need for an orderly structure is even higher. Consider the best hierarchy for files, deciding whether a deep or shallow hierarchy is preferable.

EMBRC-ERIC recommends for the EMBRC operators to utilize **relational database systems** to store those data where there is repeated need for entry, storage, querying and analysis.This implies a heavier management investment (creation and implementation of data model, administration and maintenance of relation databases, more tedious data entry, etc.), but returns higher efficiency when data need to be quality controlled, integrated, analyzed, consulted and redistributed.

Specific systems are available to support the information management of certain lab and field workflows. These LIMS (Lab Information Management Systems) and FIMS (Field Information Management Systems) have predefined data models and interfaces that can facilitate a lot of the repeating processes in the research environment.

G. Quality assurance processes

For the EMBRC research data collection, the quality control of the data can happen at different stages during the quality assurance process. An initial quality control is needed at the local level and early in the collection process. Additional control is recommended in a later stage of the data lifecycle.

The **initial quality control** of the data, **during data collection**, is the primary responsibility of the EMBRC operators (as data providers). The EMBRC operators must ensure that the recorded data reflect the actual facts, responses, observations and events. The quality of the data collection methods used strongly influences data quality, and documenting in detail how data are collected provides evidence of such quality. Errors can also occur during data entry. Data are digitized, transcribed, entered in a database or spreadsheet, or coded. Here, quality is ensured by standardized and consistent procedures for data entry with clear instructions.

Some suggestions for quality control measures during data collection or data entry:

- Calibration of instruments to check the precision, bias and/or scale of measurement;
- Taking multiple measurements, observations or samples;
- Checking the truth of the record with an expert;
- Using standardized methods and protocols for capturing observations, alongside recording forms with clear instructions
- Setting up validation rules or input masks in data entry software;
- Using data entry screens;
- Using controlled vocabularies, code lists and choice lists to minimize manual data entry;
- Detailed labelling of variable and record names to avoid confusion;
- Designing a purpose-built database structure to organize data and data files;
- Accompanying notes and documentation about the data.

(From the UK Data Archive)

An **additional quality control**, which is highly recommended for the EMBRC operators, can be performed when **checking of the data** when the data are edited, cleaned, verified, cross-checked and validated. Checking typically involves both automated and manual procedures (based on the UK Data Archive):

Some suggestions for additional quality control at data level:

- Double-checking coding of observations or responses and out-of-range values;
- Checking data completeness;
- Adding variable and value labels where appropriate;
- Verifying random samples of the digital data against the original data
- Double entry of data;
- Statistical analyses such as frequencies, means, ranges or clustering to detect errors and anomalous values;
- Correcting errors made during transcription;
- Peer review.

(From the UK Data Archive)

EMBRC-ERIC will organize training in quality assurance processes for the EMBRC operators, and this for the different types of data.

3. DOCUMENTATION AND METADATA

A crucial part of making data user-friendly, shareable and with long-lasting usability is to ensure they can be retrieved and correctly interpreted by any user. This requires clear data description, annotation, contextual information and documentation.

A. Documentation and metadata description

During the AssemblePlus project, EMBRC-ERIC will set up a digital Metadata Catalogue. This catalogue will provide discovery metadata for the EMBRC-ERIC data collection.

In order to create a digital inventory of the data resources created in the framework of EMBRC-ERIC, each EMBRC data owner will describe their generated datasets in the digital EMBRC Metadata Catalogue, using online forms.

Each data owner will make their **discovery metadata** available through the Metadata Catalogue, in order to make the data discoverable. Furthermore, each data owner will add **technical metadata** to the Meta data Catalogue, in order to interpret the data better. This includes e.g. the map projection used for geospatial resources, units of the parameters, etc. The scope, characteristics, state and accessibility of the data will be documented following common standardized formats using **metadata standards** (see further below). In case the data is already accessible through local online databases, a web link to these local systems will be included in the dataset description.

To make the collected data resources traceable and citable, EMBRC data owners have the opportunity to formally publish their datasets by the assignment of **Digital Object Identifiers (DOIs)**. The implementation of DOIs to track EMBRC resources used in scientific publications or reports can then be used to demonstrate the impact of the project.

B. Capture and creation of documentation and metadata

EMBRC-ERIC will encourage the EMBRC operators to upload their metadata in the EMBRC Metadata Catalogue as described above. Training in metadata creation is scheduled as part of the AssemblePlus project. **Online forms** for the creation of discovery and technical metadata in the EMBRC Metadata Catalogue will be provided.

EMBRC-ERIC recommends the use of LIMS and FIMS for the EMBRC operators, in order to capture data and metadata in a structured manner. A **LIMS** is a software-based Laboratory Information Management System with features that support a modern laboratory's operations. Key features include, but are not limited to, workflow and data tracking support, flexible architecture, and data exchange interfaces, which fully support its use in regulated environments. A **FIMS** or Field Information Management System enables data collection at the source (in the field) by generating spreadsheet templates, validating data, and assigning persistent identifiers to collected samples.

C. Metadata standards

EMBRC-ERIC acknowledges that common data and metadata standards and formats are a key aspect for technological and semantic data operability in order to make data discoverable for promoting international and interdisciplinary access to and use of research data. To ensure correct and proper use and interpretation of the EMBRC data by its owner and users, the use of a metadata standard is required. Standardized vocabularies and ontologies describe ways in which terms are standardized and grouped to provide consistency when ascribing metadata.

Different disciplines develop and adopt various metadata standards and/or practices for the management of their research data and materials. Some of the more commonly used metadata standards are listed below:

ISO: The International Organization for Standardization (ISO) creates documents that provide requirements, specification, guidelines or characteristics that can be used consistently to ensure that materials, products, processes and services are fit for their purpose. So far, ISO published 21.578 International Standards. A commonly used ISO standard for geographic data is the ISO 19115:2003 Geographic information -- Metadata standard, which defines how to describe geographical information and associated services, including contents, spatial-temporal purchases, data quality, access and rights to use. It is maintained by the ISO/TC 211 committee.

GCMD: NASA's Global Change Master Directory (GCMD) (https://gcmd.nasa.gov/) holds more than 34.000 Earth science data sets and service descriptions, which cover subject areas within the Earth and environmental sciences. The project mission is to assist researchers, policy makers, and the public in the discovery of and access to data, related services, and ancillary information (which includes descriptions of instruments and platforms relevant to global change and Earth science research. Within this mission, the directory also offers online authoring tools to providers of data and services, facilitating the capability to make their products available to the Earth science community. In addition, citation information to properly credit data set contributions is offered, along with direct links to data and services. As an integral part of the project, keyword vocabularies have been developed and are constantly being refined and expanded. These vocabularies are also used in other applications within the broader scientific community.

EML: Ecological Metadata Language (EML) is a metadata standard developed by and for the ecology discipline. It is based on prior work done by the Ecological Society of America and others. EML is implemented as a series of XML document types that can be used in a modular and extensible manner to document ecological data. Each EML module is designed to describe one logical part of the total metadata that should be included with any ecological dataset.

OpenAIRE (OAI-PMH v2.0): The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) provides an application-independent interoperability framework based on metadata harvesting (http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm).

4. STORAGE AND BACKUP

A. File based repositories

Repositories are needed to keep track of the raw and processed data by the users of the EMBRC services. As part of AssemblePlus, EMBRC will set up a dedicated file based repository for data which fits falls out of already existing repositories and can be used as a repository for continued storage for the data files generated from the user access projects. This is part of the tools that will be required to create DOI traceable and citable data publications, applying workflows proposed by the Publishing Data Interest Group of the Research Data Alliance (RDA) and ICSU World Data System (WDS). The dedicated repository is seen as an archival storage facility and does not exclude contributions to recognized data initiatives such as EurOBIS, EMODnet, EBI, Pangaea, GEOSS and Copernicus.

B. Assurance of adequate storage capacity

The EMBRC WGEI Strategy Report identified data storage as an e-infrastructure requirement for EMBRC. EMBRC-ERIC is committed to supply adequate storage to support the central services like the access system and repositories.

In general the issue is primarily a local one and the primary responsibility to ensure adequate storage capacity lies with the EMBRC operators. However EMBRC-ERIC engages to map the needs and look into co-operation with European initiatives to empower the local and central services. As part of the WGEI activities initial contacts have been initiated with EGI (European Grid Initiative), IndigoDataCloud and related initiatives that will take shape under the European Open Science Cloud.

C. Responsibilities for back-up and recovery

The primary responsibility for back-up and recovery of the data lies with the EMBRC operators. For the central services and repository, the central IT support for EMBRC will take up the role for back-up and recovery.

It is recommended for the EMRBC operators to create regular back-ups of the EMBRC data. Back-ups of the data should be stored at a different location than the actual data, on physically separated media. Both incremental backups (e.g. on weekdays; backups of all changed data files) and full backups (e.g. in weekends; backup of all data files) are suggested.

D. Risks and mitigations regarding data security

EMBRC-ERIC will undertake all required efforts needed to protect the data, products and services against unauthorized use. The primary responsibility to take necessary measures to ensure data security lies with the EMBRC operators.

Physical security, network security and security of computer systems and files all need to be considered to ensure security of data and prevent unauthorized access, changes to data, disclosure or destruction of data.

Some suggestions for **physical data security**:

- Controlling access to rooms and buildings where data, computers or media are held;
- Logging the removal of, and access to, media or hardcopy material in store rooms;
- Transporting sensitive data only under exceptional circumstances, even for repair purposes, e.g. giving a failed hard drive containing sensitive data to a computer manufacturer may cause a breach of security.

(From the UK Data Archive)

Some suggestions for **network security**:-

- Not storing confidential data such as those containing personal information on servers or computers connected to an external network, particularly servers that host internet services;
- Firewall protection and security-related upgrades and patches to operating systems to avoid viruses and malicious code;
- Install anti-virus packages and schedule regular scans.

(From the UK Data Archive)

Some suggestions for **security of computer systems and files**:

- Locking computer systems with a password and installing a firewall system;
- Protecting servers by power surge protection systems through line-interactive uninterruptible power supply (UPS) systems;
- Implementing password protection of, and controlled access to, data files, e.g. no access, read only, read and write or administrator-only permission;
- Controlling access to restricted materials with encryption;
- Imposing non-disclosure agreements for managers or users of confidential data;
- Not sending personal or confidential data via email or other file transfer means without first encrypting them;
- Destroying data in a consistent manner when needed;
- Remember that file sharing services such as Google Docs or Dropbox may not be that secure.

(From the UK Data Archive)

E. Assurance to secured access

EMBRC-ERIC will undertake all efforts required to provide secure access to data. Where applicable, authentication systems will be used, requesting log-in before providing access to secured data and information.

Furthermore, EMBRC-ERIC will take measures to be compliant with the EU regulations regarding the protection of personal data (http://ec.europa.eu/justice/data-protection/).

5. SELECTION AND PRESERVATION

A. Data to be retained or destroyed for contractual, legal, or regulatory purposes

In principle, all EMBRC data will be digitally preserved, unless it is stated otherwise, or the data owner gives specific reasons to not preserve the data.

For the preservation process, a distinction should be made between the **raw data**, the **intermediate data** (e.g. in between shapefiles for the creation of GIS layers) and the **final, processed data**. Raw and processed data should always be preserved. For intermediate data it might be less relevant to always preserve this data.

In order to achieve transparency about the conducted research, EMBRC **data linked to publications** should always be made accessible and retrievable.

Some reservations exist about the preservation of **medical data**. In the case of EMBRC generating patient related (medical) data, then these data cannot be preserved for the long term without anonymization. We refer to the ELIXIR Scientific Advisory Board (SAB). The ELIXIR SAB plays a major role in the process for reviewing and selecting ELIXIR Nodes and provides strategic scientific advice to the ELIXIR Board. The SAB is an independent body, made up for leading experts from around the world. EMBRC will use the recommendations on Ethical, Legal and Social Implications of the ELIXIR SAB as guidelines.

B. Foreseeable research uses for the data

The EMBRC user community covers a very wide panel of scientific fields, using approaches as diverse as molecular and cell biology, biochemistry, genomics, behavioral and reproductive biology, ecology, population genetics and host-pathogen relationships.

Possible applications for the EMBRC data are in ecological, fundamental and applied research, policy support, conservation, blue technology, pharmaceutical research, etc.

Application sectors range from gene and cell engineering (molecular farming, cell factories), bio-refineries, biostatistics, software development, nutrition, medicine and health care, aquaculture, crop disease control and environmental remediation, to bioenergy and biomaterials.

C. Long-term preservation plan

EMBRC-ERIC acknowledges the importance of long-term preservation of research data in trustworthy accredited repositories where the necessary measures are in place to archive and preserve data over long time spans. EMBRC –ERIC will make use of these repositories to comply with requirements of long term preservation. Repositories used for long term preservation should be accredited as part of recognized data initiatives like ICSU World Data System (WDS), International Oceanographic Data Exchange, etc.

6. DATA ACCESS AND SHARING

A. Data access policy

Currently, the EMBRC-ERIC Data Policy is being drafted separately.

Following Article 22 of the EMBRC-ERIC Statutes, the EMBRC-ERIC Data Policy states the EMBRC-ERIC views regarding data access, data sharing and rights. The EMBRC-ERIC Data Policy builds on the general principles described by the EMBRC-EIRC Statutes and the EMBRC Technical and Scientific Description for research data.

Objectives of the EMBRC-ERIC Data Policy are:

1. The EMBRC-ERIC Data Policy covers Data acquired, assembled or created through research, survey and monitoring activities by or involving EMBRC-ERIC that are either fully or partially funded. The EMBRC-ERIC Data Policy also applies to Data managed by EMBRC-ERIC.

2. EMBRC-ERIC will promote e-infrastructure interoperability and standardization in order to deal with large volumes of different types of generated Data, for which EMBRC-ERIC has Data handling protocols, tools and expertise in place.

3. EMBRC-ERIC acknowledges the importance of long term preservation of research Data in trustworthy accredited repositories EMBRC-ERIC will maintain a repository for ensuring archival and long term preservation of newly generated Data in collaboration with recognized Data repositories such as EurOBIS, EMODnet, Pangaea, GEOSS and Copernicus.

B. Data sharing

EMBRC-ERIC believes in the concept of **FAIR data** (Findable, Accessible, Interoperable and Re-usable) and will work towards offering EMBR research data as FAIR data (https://www.force11.org/group/fairgroup/fairprinciples).

EMBRC-ERIC will work towards providing **virtual access** to the data that is considered part of the service offer. For this a model of digital data publication is foreseen. First steps towards virtual access will be realized as part of AssemblePlus. Data that can be brought into the public domain will be made accessible by creating DOI traceable and citable data publications, applying workflows proposed by the Publishing Data Interest Group of the Research Data Alliance (RDA) and ICSU World Data System (WDS). In order to create a digital inventory of the data resources available at the AssemblePlus stations, each of the partners will describe their datasets in a digital catalogue, using an online form. Data that can be brought into the public domain will be prepared for **data publication**. A technical quality control will verify that all required information is included. Each of the datasets proposed for publication will be made citable and will be labeled with a **Digital Object Identifier (DOI)**.

To stimulate discovery and potential re-use of the data, AssemblePlus will maximally exploit existing data flow pathways to share data through European e-infrastructures like EMODnet, LifeWatch and ELIXIR and global initiatives like the Global Biodiversity Information Facility (GBIF) and the Global Earth Observation System of Systems (GEOSS), especially its biodiversity component (GEOBON).

EMBRC-ERIC promotes a culture of openness and sharing of data and will therefore stimulate the exchange of good practices in data access and sharing by liaising with existing European initiatives or relevance for environmental and biological data and bioinformatics.

Annex 1 lists some potential (European and non-European) data repositories for redistribution for the thematic data types generated within EMBRC-ERIC.

7. REPONSIBILITIES AND RESOURCES

A. Responsibilities for implementing and actualizing the DMP

Implementation of the DMP is necessary at both the central and local level. The final responsibility for implementing the DMP lies with EMBRC-ERIC. The EMBRC operators will ensure the DMP is being implemented at the local level. EMBRC-ERIC will support the operators by organizing the necessary training and the establishment of a data management working group. This working group of data scientists will have a representative for each EMBRC operator. This working group could also take responsibility for the actualization of the DMP where needed.

B. Centralized and distributed responsibilities and roles of partners

Throughout this EMBRC-ERIC DMP, the responsibilities for several steps in the data management process are described. They are listed again below.

EMBRC-ERIC will advise and train its members to organize a structured data storage.

The initial quality control of the data, during data collection, is the primary responsibility of the EMBRC operators (as data providers). Additional quality controls, during checking of the data are highly recommended for the EMBRC operators. EMBRC-ERIC will organize training in quality assurance processes for the EMBRC operators, and this for the different types of data.

In order to create a digital inventory of the data resources created in the framework of EMBRC-ERIC, each EMBRC data owner will describe their generated datasets in the digital EMBRC Metadata Catalogue, using online forms. Each data owner will make their discovery metadata available through the Metadata Catalogue, in order to make the data discoverable. Furthermore, each data owner will add technical metadata to the Meta data Catalogue, in order to interpret the data better. EMBRC-ERIC will encourage the EMBRC operators to upload their metadata in the EMBRC Metadata Catalogue as described above.

In general the issue is primarily a local one and the primary responsibility to ensure adequate storage capacity lies with the EMBRC operators. However EMBRC-ERIC engages to map the needs and look into co-operation with European initiatives to empower the local and central services.

The primary responsibility for back-up and recovery of the data lies with the EMBRC operators. For the central services and repository, the central IT support for EMBRC will take up the role for back-up and recovery.

EMBRC-ERIC will undertake all required efforts needed to protect the data, products and services against unauthorized use. The primary responsibility to take necessary measures to ensure data security lies with the EMBRC operators.

EMBRC-ERIC will undertake all efforts required to provide secure access to data. Where applicable, authentication systems will be used, requesting log-in before providing access to secured data and information.

Furthermore, EMBRC-ERIC will take measures to be compliant with the EU regulations regarding the protection of personal data (http://ec.europa.eu/justice/data-protection/).

EMBRC-ERIC will work towards providing virtual access to the data that is considered part of the service offer.

EMBRC-ERIC promotes a culture of openness and sharing of data and will therefore stimulate the exchange of good practices in data access and sharing by liaising with existing European initiatives or relevance for environmental and biological data and bioinformatics.

C. Centralized and distributed resource requirements for implementation

Implementation of the DMP is necessary at both the central and local level. The EMBRC operators will ensure the DMP is being implemented at the local level. For implementation at the central level, a data management working group will be established. This working group of data scientists will have a representative for each EMBRC operator.

Furthermore, resources are needed to support the services discovery and request system, the metadata catalogue and the data repository.

D. Requirements of specialist expertise and equipment

A data scientists or custodians should be appointed for each EMBRC operator and be the representative for the EMBRC operator as part of the data management working group. Data management capacity and expertise is also needed to support the central services and guide the implementation of the EMBRC-ERIC DMP. IT capacity is needed to set up and maintain the central systems.

ANNEX 1: Potential (European and non-European) data repositories for redistribution of the thematic data types within EMBRC-ERIC

	Thematic Data type category	Research domains	Potential data repositories for redistribution	URLs
Biological data	Biodiversity data	e.g. biogeographic research, species traits, taxonomy	European Ocean Biogeographic Information System (EurOBIS)	http://www.eurobis.org/
			Ocean Biogeographic Information System (OBIS)	http://www.iobis.org/
			Global Biodiversity Information Facility (GBIF)	http://www.gbif.org/
			Biology Portal of the European Marine Observation and Data Network (EMODnet Biology)	http://www.emodnet.eu/biology
			World Register of Marine Species (WoRMS)	http://www.marinespecies.org/
			Morphobank.org	https://morphobank.org//
	Genomic data	e.g. Sequencing (DNA, RNA,), annotation of features, protein structural information, gene expression profiles, alignment data, chromosomal mapping, phylogenetic trees, Single Nucleotide Polymorphisms (SNPs), functional genomics, Metabolomics, Proteomics, environmental DNA	European Nucleotide Archive (ENA)	http://www.ebi.ac.uk/ena
			GenBank	https://www.ncbi.nlm.nih.gov/genbank/
			DNA Data Bank of Japan (DDBJ)	http://www.ddbj.nig.ac.jp/
			dbSNP	https://www.ncbi.nlm.nih.gov/snp
			European Variation Archive (EVA)	http://www.ebi.ac.uk/eva/
			dbVar	https://www.ncbi.nlm.nih.gov/dbvar/
			Database of Genomic Variants Archive (DGVa)	http://www.ebi.ac.uk/dgva
			EBI Metagenomics	https://www.ebi.ac.uk/metagenomics/
			NCBI Trace Archive	https://www.ncbi.nlm.nih.gov/Traces/home/
			NCBI Sequence Read Archive (SRA)	https://www.ncbi.nlm.nih.gov/sra
			NCBI Assembly	https://www.ncbi.nlm.nih.gov/assembly

			NCBI Reference Sequence Database (RefSeq)	https://www.ncbi.nlm.nih.gov/refseq/
			Entrez Nucleotide	https://www.ncbi.nlm.nih.gov/nucleotide/
			UniGene	https://www.ncbi.nlm.nih.gov/unigene
			HomoloGene	https://www.ncbi.nlm.nih.gov/homologene/
			Protein Information Resource (PIR)	http://pir.georgetown.edu/
			Protein Data Bank (PDB)	https://www.wwpdb.org/
			UniProt	http://www.uniprot.org/
			Entrez Protein	https://www.ncbi.nlm.nih.gov/protein
			Protein Circular Dichroism Data Bank (PCDDB)	http://pcddb.cryst.bbk.ac.uk/home.php
			ArrayExpress	http://www.ebi.ac.uk/arrayexpress/
			Gene Expression Omnibus (GEO)	https://www.ncbi.nlm.nih.gov/geo/
			GenomeRNAi	http://www.genomernai.org/
			dbGAP	https://www.ncbi.nlm.nih.gov/gap
			The European Genome-phenome Archive (EGA)	https://www.ebi.ac.uk/ega/
			Database of Interacting Proteins (DIP)	http://dip.doe-mbi.ucla.edu/dip/Main.cgi
			IntAct	http://www.ebi.ac.uk/intact/
			Japanese Genotype-phenotype Archive (JGA)	http://trace.ddbj.nig.ac.jp/jga/index_e.html
			Biological General Repository for Interaction Datasets	https://thebiogrid.org/
			NCBI PubChem BioAssay	https://pubchem.ncbi.nlm.nih.gov/
			MetaboLights	http://www.ebi.ac.uk/metabolights/
			PeptideAtlas	http://www.peptideatlas.org/
			PRIDE	http://www.ebi.ac.uk/pride/archive/

			ProteomeXchange	http://www.proteomexchange.org/
	Imaging data	eg. Zooscan images, flowcam images, flow cytometry images, VPR videos	EcoTaxa	http://ecotaxa.obs-vlfr.fr/
			FlowRepository	http://flowrepository.org/
	Biogeochemical data	e.g. Biochemical pathways, Nutrients	Data Exchange Portal of MPI	https://www.bgc-jena.mpg.de/geodb/projects/Home.php
	Experimental data	e.g. Data resulting from lab experiments		
Oceanographical data	**Physical data**	e.g. Sea Water Temperature, Salinity, ocean currents, waves	The Physical Portal of the European Marine Observation and Data Network (EMODnet Physics)	http://www.emodnet-physics.eu/map/
	Chemical data	e.g. Pollution, Heavy metals	The Chemical Portal of the European Marine Observation and Data Network (EMODnet Chemistry)	http://www.emodnet-chemistry.eu/
Climatological data	**Meteorological data**	e.g. Air temperature, Wind speed, …		
	Modelling data	e.g. Climate models	BioModels Database	http://www.ebi.ac.uk/biomodels-main/
			Kinetic Models of Biological Systems (KiMoSys)	https://kimosys.org/

ANNEX 2: Virtual tool-box of best practice guidelines

Assemble "virtual tool-box of best practice guide-lines" (http://www.assemblemarine.org/virtual-tool-box-of-best-practice-guide-lines/). This toolbox aggregates the following protocols and guidelines:

- On-site access – hosting guest scientists
    - Draft Letter of Acceptance
    - Best practice guidelines for on-site access
- Remote access – shipping of marine organisms
    - Best practice guidelines for remote access
- Whole, multicellular marine organisms – holding, breeding, culturing, etc.
    - Sampling strategy for *Ciona intestinalis*
    - Fertilization tests in *Ciona intestinalis*
    - Material and packaging techniques for the shipment of *Ciona intestinalis*
    - Health and maturation status of *Ciona intestinalis*
    - Axenic *Ectocarpus* cultures
    - Biolistic delivery to *Ectocarpus* cells
    - Genetic crosses in *Ectocarpus*
    - Extraction of total protein from *Ectocarpus*
    - Isolation and regeneration of protoplasts from *Ectocarpus*
    - *Ectocarpus* RNA extraction method
    - Preparation of seawater media for *Ectocarpus* culture
    - How to cultivate *Ectocarpus*
    - Immunostaining of *Ectocarpus* cells
    - Isolation of *Ectocarpus* gametophytes
    - Passage of *Ectocarpus* cultures
    - Spawning of gametes in *Paracentrotus lividus*
    - Induced spawning of gametes in *Ciona intestinalis*
    - Fertilization in *Paracentrotus lividus*
    - Dissection of the endostyle in *Ciona intestinalis*
    - Improved maintenance protocols for corals (*Stylophora pistillata*)
    - Culture techniques for seagrasses
    - Maintenance of *Amphioxus* and induction of spawning
    - Improved maintenance protocol for corals (*Lophelia pertusa*)
    - Culture technique for Hagfish
    - Culture techniques for *Fucus vesiculosus* or *F. serratus*
    - Improved maintenance protocols for kelp gametophytes
- Marine protists & cell lines of marine animals - development, transfection, cryopreservation, etc.
    - Primary cultures of functional Atlantic cod melanophores

- Detection of glycosaminoglycans and fibrous collagen in marine fish cell lines
- Establishing non-axenic monoclonal cultures of *Skeletonema marinoi*
- Detection of mineral nodules in marine fish mineralogenic cell lines
- Development of primary cell cultures from calcified tissues of marine fish
- Nucleofection of marine fish cell lines
- Development of primary cell cultures from *Ciona intestinalis* explant
- Preparation of fish serum for the culture of marine fish cell lines
- Cryopreservation and revival of marine fish cell lines
- Detection of alkaline phosphatase (ALP) activity in marine fish cell lines
- Development of primary cell cultures from *Ciona intestinalis* larval stage
- Transfection of marine fish cell lines using polyethylenimine (PEI)
- Semi-quantitative analysis of cell proliferation in *Asterias rubens* cell monolayers
- Phagocytic behavior of *Asterias rubens* blood cells or coelomic epithelia cells in vitro
- Epithelial cell primary cultures isolated from buds of *Botryllus schlosser*
- Preparation of Ascidian hemolymph for the culture of *Ciona* cells
- Preparation of sea urchin gametes and development of sea urchin embryos
- Primary cell culture from pituitary of *Dicentrarchus labrax*
- Primary cell cultures from total *Paracentrotus lividus* coelomocytes
- Cryopreservation of marine microalgae employing a controlled rate cooler
- Cryopreservation of *Ectocarpus*
- Cryopreservation of marine microalgae employing a passive freezer
- Automated extraction of PCR-grade gDNA from *Ciona intestinalis* tissue
- Automated whole-mount in situ hybridization on developmental stages of *Ciona intestinalis* for the identification of reversive mutations with subtle phenotype
- Cryopreservation protocol for *Ciona intestinalis* sperm
- Screening protocol for the identification of spontaneous mutations in *Ciona intestinalis*
- Maintenance of marine amoebae
- Maintenance of marine ciliates
- Isolation of algal endosymbionts from protists
- Maintenance of marine heterotrophic protists
- Use of FlowCam to enumerate and differentiate between marine protists
- Elimination of bacteria from microalgal culture using antibiotics
- Tips for physical isolation of bacteria-free, clonal microalgae from marine environmental samples
- Medium scale culture of micro-algae in polycarbonate carboys
- Medium scale culture of micro-algae in plastic bags
- AFLP analysis as a tool to investigate genetic diversity in microalgae
- Molecular barcoding of protists
- Culturing of *Pseudo-nitzschia* species on agar medium
- Diatom cleaning with nitric/sulfuric acids
- Dinoflagellate isolation from Cnidarian
- Microalgae preparation for scanning electron microscopy; dehydration

- Genetic and genomic resources
    - Extraction of high quality genomic DNA from *Ectocarpus*
    - Preparation of plugs of *Ectocarpus* material for pulse field electrophoresis
    - UV mutagenesis of *Ectocarpus* gametes
    - *Phaeodactylum tricornutum* cryopreservation
    - *Phaeodactylum tricornutum* transformation by means of the PDS-1000/He Microprojectile Accelerator