



# Operational Sensing Life Technologies for Marine Ecosystems

## Deliverable D1.3 – Data Management Plan

**Lead Beneficiary:** Pensoft Publishers

**Author/s:** Gabriela Popova, Nikola Ganchev

21/06/2023



Funded by  
the European Union

Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the EU nor the EC can be held responsible for them.

**Prepared under contract from the European Commission**

Grant agreement No. 1010949

EU Horizon Europe Research and Innovation action

Project acronym: **ANERIS**  
Project full title: **operAtional seNsing lifE technologies for maRIne ecosystemS**  
Start of the project: January 2023  
Duration: 48 months  
Project coordinator: Jaume Piera

Deliverable title: Data Management Plan  
Deliverable n°: D1.3  
Nature of the deliverable: Report  
Dissemination level: Public

WP responsible: WP1  
Lead beneficiary: PENSOFT

Citation: Popova, G., & Ganchev, N. (2024). *Data Management Plan*.  
Deliverable D1.3 EU Horizon Europe  
ANERIS Project, Grant agreement No. 101094924

Due date of deliverable: Month n°6  
Actual submission date: Month n°6

Deliverable status:

Version	Status	Date	Author(s)
1.0	Draft	21 June 2023	Gabriela Popova, Nikola Ganchev PENSOFT
1.1	Review	27 June 2023	Anna Sapundzhieva, Pavel Stoev PENSOFT
1.2	Draft	27 June 2023	Gabriela Popova, Nikola Ganchev PENSOFT
1.3	Review	29 June 2023	Alina Luna, Jaume Piera CSIC

The content of this deliverable does not necessarily reflect the official opinions of the European Commission or other institutions of the European Union.

## Table of Contents

<b>Preface</b> .....	<b>4</b>
<b>Summary</b> .....	<b>4</b>
<b>List of Abbreviations</b> .....	<b>5</b>
<b>1. Introduction</b> .....	<b>5</b>
<b>2. Data Summary</b> .....	<b>6</b>
<b>3. FAIR Data</b> .....	<b>15</b>
3.1 Making data findable.....	15
3.2 Making data openly accessible.....	16
3.3 Making data interoperable.....	18
3.4 Increasing data re-use.....	19
<b>4. Other research outputs</b> .....	<b>19</b>
<b>5. Allocation of costs</b> .....	<b>19</b>
<b>6. Data security</b> .....	<b>20</b>
<b>7. Intellectual property</b> .....	<b>20</b>
<b>8. Ethics</b> .....	<b>21</b>
<b>References</b> .....	<b>21</b>
<b>Annex</b> .....	<b>21</b>

## Preface

ANERIS, a research project funded by the European Commission's Horizon Europe programme, will adhere to all regulations governing Open Science, Open Research Data, and Research. By effectively organising, securely storing, and ensuring accessibility of research data, ANERIS aims to continuously monitor their validity, resulting in efficient and high-quality research outcomes and cost savings. The consortium behind ANERIS is committed to advancing open science policies and practices by ensuring that research data is Findable, Accessible, Interoperable, and Reusable (FAIR). To achieve this, a project-specific D1.3 Data Management Plan (DMP) has been devised, guaranteeing resilient and transparent management of research data generated and utilized within ANERIS.

As this deliverable is submitted during the early stage of the project (M6), it addresses data management questions relevant to this specific phase. While certain details may not be fully covered, the DMP is a living document that will undergo at least one update throughout the implementation of ANERIS.

The development of the DMP has involved close collaboration with project partners who provided information regarding their planned usage and generation of research data, as well as their institutional policies on data management. This information was collected through a questionnaire distributed during M5 of the project, utilising LimeSurvey, an open-source online statistical survey platform. Annex 1 of this document includes the questionnaire for reference.

## Summary

The data management plan of ANERIS strives to establish guidelines and suggestions for the availability of data in a FAIR manner, covering the entire project lifespan and extending beyond. This plan governs the handling of data both within ANERIS's research network, encompassing data generation, and the utilisation and processing of primary and secondary data obtained from sources external to ANERIS, such as external datasets.

In line with that, this DMP adheres to a well-defined structure, presenting a Data summary that encompasses the datasets ANERIS intends to utilise and generate. This summary includes details such as the purpose, format, origin, expected size, utility, and curation of the datasets (Chapter 1). Following the Data summary, there are dedicated sections outlining ANERIS's strategies for ensuring data findability, accessibility, interoperability, and reusability (Chapter 2). The DMP will also outline the allocation of resources for FAIR data management within the consortium, both during the project's lifespan and beyond (Chapter 3). Additionally, the DMP will address data security practices, ensuring that necessary measures are in place to preserve and curate research data (Chapter 4). Lastly, Chapter 5 delves into the ethical considerations of data sharing, including compliance with GDPR regulations when handling personal data.

## List of Abbreviations

DMP	Data Management Plan
EU	European Union
DPO	Data Protection Officer
FAIR	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
WP	Work Package
DOI	Digital Object Identifier
IP	Intellectual Property

## 1. Introduction

As a Horizon Europe project, ANERIS upholds the fundamental principles of open science, fostering collaboration, utilising open tools, and promoting the dissemination of knowledge. By embracing these principles, ANERIS aims to transcend the boundaries of the project itself, ensuring that its research outcomes and significant discoveries have a lasting impact on research progress and society as a whole. To achieve this, the project places great importance on equipping its partners with well-defined procedures and guidelines that enable the provision of open access to research outputs without any access charges or barriers, facilitating widespread engagement among potential end-users. Moreover, to enhance the seamless exchange of data, the project's website contains a secure partner area for project management. The portal allows the sharing and archiving of electronic conversations between partners, document collaboration features, a repository for data management, and project tracking.

In alignment with this objective, ANERIS's work package 1, undertakes the responsibility of formulating D1.3 Data Management Plan (DMP) within the initial six months of the project. The DMP is being developed for the following key purposes: 1) comprehensively documenting the utilisation of data throughout ANERIS's scientific endeavors, 2) facilitating future implementation of research data and its derivatives for further exploration, supported by machine-readable metadata, and 3) ensuring robust data security measures, protection, and documentation. The DMP consolidates ANERIS's data management practices, elucidating the data types, licenses, and formats to be employed in adherence to the FAIR data principles. Furthermore, D1.3 encompasses a set of recommendations that guide project members in the generation, collection, and utilisation of research data, taking into account ethical considerations such as the proper handling of sensitive or personal data. The DMP delineates the datasets earmarked for open usage within the project, while also identifying trusted data repositories where they can potentially be shared. To enhance the value of the DMP, a concise supplementary document containing Data Management Guidelines will be produced and shared with partners, succinctly summarising the current recommendations and serving as a practical reference.

The development of the DMP is rooted in the initial open access principles outlined in the project's description of action, supplemented by valuable insights gathered through an inclusive

consultation process involving the consortium members. Project partners actively participated in a comprehensive survey, consisting of 13 in-depth questions (available in Annex 1), which extensively covered aspects such as the discoverability, accessibility, interoperability, and reusability of ANERIS's research results and data. The survey was diligently completed by the 22 out of the 25 project organisations responsible for storing, generating, or reusing data pertinent to ANERIS's research endeavors, namely ICM, CSIC, Quanta, CytoBuoy, MarsBased, FECDAS, CNRS, VLIZ, HCMR, Euro-Bioimaging, BIOPOLIS, Sorbonne university, EMSO ERIC, Marine Institute - Foros na Mara, UPC, LifeWatch ERIC, EMBL, OSLOMET, MedCities, NORCE and University of Haifa. Each of the partnering institutions provided accurate responses to the survey, encompassing details about the datasets they will generate and/or reuse, as well as their institutional practices and preferences concerning data management.

## 2. Data Summary

ANERIS will collect, generate, analyse and reuse data in order to conceptualize, create and validate the 11 planned ANERIS technologies and achieve the overarching case studies and OMB goals:

- **(CS1)** Improving the capabilities of RI observatories for continuous collection, sorting and analysis of high-temporal resolution marine life data. Aside from the engineering aspects of this work, cytometric (.csv) and image data (.jpeg, .jpg, .TIFF) of plankton and other marine life will be generated from the observatories. Sorting and analysis of this data will be achieved partly via automated classifiers and methodologies developed during the project with the aim of scaling up FAIR data management and dissemination pipelines.
- **(CS2)** Improving the spatial and temporal resolution of marine life monitoring based on genomics. The technologies related to this case study will extend the capabilities and scope of the existing network of genomic-based monitoring through developing and optimising protocols for nanopore sequencing, increasing the accessibility and interoperability genomic data (.fasta, .fast5, .fastq) through the use of DwC-A and MIXS data standards for published data and improving accessibility of software for genetic analysis.
- **(CS3)** Large scale marine participatory actions (bioblitzes) will engage scientists and citizens around the European seas and involve them in collecting vast amounts of image data (.jpeg, .jpg, .TIFF) of shallow- and deep-water marine life. These images will serve as training data for the multiple information-extraction, image restoration, image extraction and hybrid-intelligence species identification technologies which will significantly improve analysis automation and reliability. This will be supplemented with an accessible web interface for uploading data thus improving the FAIR collection, processing and management of marine data.

- **(CS4)** Merging imaging and genomic information in different monitoring scenarios will aim to increase both the taxonomic resolution and the accuracy of marine biodiversity data by utilizing algorithms and AI-approaches for simultaneous analysis of genomic (.csv) and imaging (.jpeg, .jpg, .TIFF) data.

The combination of these goals aims to result in the operationalization, validation and adoption of OMB methods and scaling up the FAIR data management and dissemination pipelines using project-generated AI-driven analysis and publishing (EGI, EOSC). To achieve this, ANERIS will primarily use self-generated data from observatories and RI's operated by project partners, as well as from direct observations collected from field work and large-scale citizen science (e.g bioblitzes). A preliminary overview of the planned datasets that will be generated can be found on Table 1. The majority of the datasets will be open access with suitable ones also uploaded to global biodiversity databases and repositories (e.g GBIF, EurOBIS), in Zenodo or as data papers in e.g RIO journal, complying with the policy and recommendations on reproducibility of scientific results in the EU. Few datasets (like No 2, 7, 9 & 10 on Table 1) will be closed access due to containing personal or sensitive information subject to GDPR regulations ((EU 2016/67).

Datasets will contain both quantitative and qualitative data concerning primarily environmental data (images, genomic data) with varying geographic, MIXS, DwC-A, or other conventional biodiversity metadata. These datasets significantly vary in size but are generally multiple hundred GBs and up to several TBs. The project will also collect personal data in restricted-access comma-separated values (.csv) files from citizen science or workshop activities, including names, phone numbers and email addresses of participants. The anticipated size of these datasets is generally under 1MB.

ANERIS will also reuse data from different formats, e.g. .jpg, .jpeg, .TIFF, .csv, .fasta, .fast5, .fastq, xlsx (Table 2). These data primarily originate from previous projects (cos4life, BioMARato 2021, Jericho-Next 2019, BG-PART, EMSO Sensor arrays, EMOBON, SmartBay, etc) or data previously generated by partner institutions and RIs (University of Haifa training datasets & EMSO environmental data).

Aside from contributing to the objectives of ANERIS and the specific needs of its work packages, case studies and technologies, the generated data will also be of use to the following stakeholder groups (more information in *D6.1 Plan for Communication and Dissemination of Results (PCD)*):

- Scientific community
- Policymakers and advisors
- Citizen scientists and citizen science networks
- The general public

An initial summary of the generated and reused data for ANERIS can be found in Table 1 and Table 2, respectively. These tables were created using partners' responses to the DMP survey that can be found in Annex 1. Since not all information regarding the generated and reused

datasets is available at this early project stage, the DMP will be treated as a living document which will be updated with necessary amendments and information until it fully reflects the landscape of ANERIS research data.



**Table 1. Summary of the anticipated data that will be generated by ANERIS (More info about Licences and Points of Access can be found in Section 3)**

No	Name of the dataset	Name of the generator	Relevant task	Generated via	Size	Format	Type of data	Sensitive personal data	Personal data	Delivery	Type of metadata	Users	Access
1	Inputs from the co-design events, co-design digests	Science for Change SL	Collect the insights, information and details from the co-design sessions (T2.1; T3.1; T4.1)	Co-design workshops	TBD	.pdf, .docx, .xlsx, Miro board, photos	Qualitative	No	Yes	15 days to 1 month after the session	DOI, filters	Researchers, policy makers, communities	Open access
2	Registered participants in co-design workshops	Science for Change SL	List and track the invitations to co-design events (T2.1; T3.1; T4.1)	Partners, desk research, networking etc.	>100MB	.xlsx	Quantitative and qualitative	Yes	Yes	7-15 days before the event	Filters	Internal	Closed access
3	Post-event surveys	Science for Change SL	Evaluate the co-design sessions (T2.1; T3.1; T4.1)	Survey	>100MB	.xlsx	Quantitative and qualitative	Yes	Yes	15 days to 1 month after the session	Filters	Researchers	Open access
4	Source code	Quanta Systems, S.L.	Software development (WP4; WP5)	Developers + co-design	MBs	Text format	Source code	No	No	In each deliverable	Documentation via MD files	Developers	Open access
5	measurement_date+time	CytoBuoy	Generate scan and images of particles (T3.4)	In-situ measurement	< 1GB	.cyz	Quantitative	No	No	From the moment the instrument is installed at OBSEA and SMARTBAY	image and flow definition set, .cyz definition file and target ranges	Researchers, Phd, contract researchers	Open
6	Source code	MarsBased	Software development (WP4; WP5)	Developers + co-design	MBs	Text format	Source code	No	No	Released with each deliverable	Documentation via MD files	Developers	Open access

D1.3 Data Management Plan  
ANERIS #101094924

7	ANERIS' Activities Inscriptions	FECDDAS	Signing up for ANERIS activities (T4.6; T6.2)	Survey or form (Google Forms)	>1MB	.xlsx	Qualitative	Yes, health data.	Yes	None	TBD	ANERIS team	Closed access, for privacy reasons.
8	ANERIS' Online Webinars	FECDDAS	Signing up for ANERIS online webinars (T4.6; T6.2)	Survey or form (MailChimp)	>1MB	.xlsx	Qualitative	No.	Yes	None	TBD	ANERIS team	Closed access, for privacy reasons.
9	ANERIS' Volunteers Information	FECDDAS	Database of volunteers for engaging with ANERIS activities (T4.6; T6.2)	Data aggregation.	>1MB	.csv	Qualitative	No.	Yes	None	TBD	ANERIS team	Closed access, for privacy reasons.
10	ANERIS' Results	FECDDAS	Number of volunteers, diving locations, photos taken (T4.6; T6.2)	Data aggregation.	>1MB	.xlsx	Qualitative and quantitative	No.	Yes.	None	TBD	ANERIS team	Closed access, for privacy reasons.
11	Validation annotated data for AEIS-MAC	CNRS	T3.6	data aggregation, discussion and training with experts	~1GB	COCO Dataset JSON Format	quantitative Annotations of datasets provided in T3.1 and T3.5	No	No	M20	use of the standard of COCO	developers	open access
12	Extracted parameters by AEIS-MAC that can feed the characterisation and the taxonomic classification on validation data	CNRS	T3.6; T4.4	AEIS-MAC analysis of datasets	TBD	JSON	Quantitative extracted information from dataset	No	No	M24	Standardized description of processing for each index	4.4, public databases such as ecoTAXa	open access
13	Micro(phyto)plankton observations through imaging	VLIZ	T3.4	Imaging FlowCytometry (CytoSense)	GB	.csv TIFF	Qualitative Quantitative	No	No	By the end of the ANERIS project	Biodiversity standards. Metadata discovery	Quadruple helix	Open access

D1.3 Data Management Plan  
ANERIS #101094924

	flowcytometry in the North Sea region from 2023-2024										record in the IMIS		
14	Single engagement dataset	ICM-CSIC	Co-design	Registration forms in engagement activities related to co-design	TBD	.xlsx	Qualitative	No	Yes	Not shared	TBD	Co-Design team in Barcelona	Closed access
15	Biodiversity MINKA dataset	ICM-CSIC	MINKA curator data	Citizen science participation, internal ANERIS group contributions	TBD	CSV	Quantitative and Qualitative	No	Yes	Available instantly	Number, time, latitude, longitude, taxon name, image url, sound url	Scientists, policy makers, naturalists, etc	Open access
16	Time series of particles concentration	CNRS/SU	T5.3	UVP deployments at Spanish and Irish sites	Raw: 10GB, Processed: 1GB	Text files, .tsv	quantitative	No	No	M29	Title, instrument, date, time, depth, lat, lon	Biogeochemists	Open after delay
17	Time series of planktonic objects concentration	CNRS/SU	T5.3	UVP deployments at Spanish and Irish sites	Raw: 20GB, Processed: 1GB	Raw: images, Processed: .tsv	quantitative	No	Yes	M31	Title, lat, lon, instrument, date, time, depth,	Plankton ecologists	Open after delay
18	Oxford Nanopore sequencing data from the Belgian Part of the North Sea (core VLIZ dataset).	VLIZ	T2.2	Oxford Nanopore sequencing and processing of raw sequence data	TBs	.csv .fasta and fastq fast5	Qualitative	No	No	Summer 2025	Biodiversity standards. Metadata discovery record in the IMIS	Quadruple helix, mainly scientists	Open access
19	DNA sequence data	HCMR /BIOPOLIS	T2.2.1	DNA sequencing	~2TB	various (incl. fast5, fastq fasta)	Quantitative	No	No	M36	MixS and DWC-A standards	Scientists	Open access

D1.3 Data Management Plan  
ANERIS #101094924

20	A for image analysis training	Euro-Bioim aging ERIC	T3.7	workshop	500 mb	mp4	Qualitative	No	No	M10	TBD	Consortium, Euro-Bioim aging Community	open access
21	Ongoing technology development	Euro-Bioim aging ERIC	T3.7	workshop	500 mb	mp4	Qualitative	No	No	M10	TBD	Consortium, Euro-Bioim aging Community	open access
22	Developed imaging tools and image analysis	Euro-Bioim aging ERIC	T3.7	workshop	500 mb	mp4	Qualitative	No	No	M10	TBD	Euro-Bioim aging Community, general public	open access
23	Workshop presentations	Euro-Bioim aging ERIC	T3.7	workshop	500 mb	PDF	Qualitative	No	No	M10, M22, M48	TBD	Consortium, Euro-Bioim aging Community, General public	open access
24	Occurrence of atlantic intertidal species	BIOPOLIS	WP4	photos	100GB	.jpeg	quantitative	No	Yes	Ongoing	coordinates, user, date, time	General public; Scientists	Open access
25	Photoquadrats	BIOPOLIS	WP3	photos	200GB	.jpeg, txt	quantitative	No	No	Ongoing	coordinates, user, date, time, slope, orientation	Scientists	Open access
26	Citizen science events registration	MedCities	T4.6	google forms	>100MB	.xlsx	contact data	No	Yes	Not shared	n/a	MedCities internal use	Closed
27	Multi view images	University of Haifa	T3.5	cameras	500GB	raw images	images	No	No	M12, M24, M36, M48	location	TBD	open

**Table 2. Summary of the anticipated data that ANERIS expects to reuse**

No	Name of dataset	Relevant task	Size	Format	Sensitive Personal data	Personal data	Metadata	Access	Origin	Ownership	Licence
1	ANERIS Results	This database is aimed to aggregate volunteer diving centers information, data about diving sites, number of participants, dates, photos taken, etc.	>1MB	.xlsx	No	Yes	TBD	Restricted	ANERIS Results	FECNAS	TBD
2	Training test and validation dataset	T3.5; T3.6	~300GB	.jpg	No	No	Provenance	Open	UH, cos4life bioblitz (BioMARato 2021)	Unknown	TBD
3	Micro(phyto)plankton biodiversity data from a LifeWatch/Jerico North Sea Cruise with R/V Simon Stevin in May 2019	Task 3.4 AIES-PHY development (VLIZ, Cytobouy, M1-48).	TBD	.csv TIFF	No	No	sample, sample processing, classification and processing metadata	Open	Generated under Jerico-Next 2019	VLIZ	CC-BY
4	Micro(phyto)plankton biodiversity data from from Greenland and the North Sea region collected under BG-PART from 2022-2024	Task 3.4 AIES-PHY development (VLIZ, Cytobouy, M1-48).	TBD	.csv TIFF	No	No	sample, sample processing, classification and processing metadata	Open	Generated under BG-PART, PhD Jens Dujardin	VLIZ, UGent	CC-BY
5	GBIF Database	Global Repository	TBD	.csv	No	Yes.	Date, location, identification, origin of the data	Open	Citizen science; research & museum contributions	Global data records	CC
6	Environmental data collected at the EMSO sites of deployment	T5.3	~2GB	NetCDF or text tables	No	No	Date, time, lat, lon	Open <a href="https://data.emso.eu">https://data.emso.eu</a>	EMSO Sensor arrays	EMSO (Public)	CCBY

D1.3 Data Management Plan  
 ANERIS #101094924

7	EMOBON	T2.2.1	1TB	fastq	No	No	MixS	Open	EMOBON marine stations	EMBRC-ERIC	CCBY
8	Multiview images	T3.5	500GB	Raw images	TBD	TBD	TBD	Open	TBD	TBD	Open
9	In situ observatory sensor Data	Case Study 1 Technology 4	1-20GB	.txt .csv	No	No	Common metadata format	Open	Sensor data from SmartBay	MI	CC-BY 4.0

## 3. FAIR Data

### 3.1 Making data findable

ANERIS will ensure that its publicly available finalised datasets are easily discoverable by assigning them globally unique and persistent identifiers, such as Digital Object Identifiers (DOIs). These identifiers will be among the key components included in the comprehensive metadata of the project datasets. Metadata refers to standardised and structured information describing various characteristics of the dataset, including its origin, purpose, time, geographic location, creator, access terms, and usage terms. The presence of metadata facilitates resource location and provides searchable information, enabling users to locate and cite existing data with ease. ANERIS is committed to providing thorough metadata descriptions for its data, explicitly incorporating the data identifier, offering descriptive and structural information, and providing a comprehensive representation of the dataset.

When generating metadata, ANERIS aims to adhere to a unified metadata description standard that aligns with the principles of data being findable, accessible, interoperable, and reusable. It is crucial to select a standard that suits the specific data type, ensuring interoperability with other datasets in the field and enhancing its discoverability. Given that ANERIS will produce various types of data, the project has identified several suitable standards that can be employed based on the nature of the data.

An exemplary structure of the minimum characteristics of metadata is proposed below:

- Author(s)
- Year
- Dataset Title
- Data Repository or Archive
- Global Persistent Identifier
- Version, or Subset, and/or Access Date
- Language
- Metadata language
- License of use
- Date of metadata creation
- Hierarchy level
- Character encoding
- Format version
- Keywords (if possible)

### Recommendations

ANERIS partners are encouraged to apply one of the following metadata standards:

- [ISO 19115](#): provides information about the identification, extent, quality, spatial and

temporal aspects, content, spatial reference, portrayal, distribution, and other properties of digital geographic data and services;

- [ISO/IEC 19506](#): suitable for representing existing software assets, their associations, and operational environments;
- [INSPIRE metadata](#): requirements for the creation and maintenance of metadata for spatial data sets, spatial data set series and spatial data services set by the EC.

Considering that the project is going to implement AI-models, which will co-develop and deploy AI-driven data analysis to support the FAIR data management and dissemination pipelines of the project, ANERIS will refer, where possible, to the “Regulation of the European Parliament and of the Council: Laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts”.

Moreover, ANERIS strives to facilitate information sharing among its consortium members, which is why it will adhere to a dedicated data naming convention. By adopting a standardised and intuitive approach to naming documents, ANERIS ensures efficient collaboration and enables partners to effortlessly locate project datasets whenever needed.

### Recommendation

ANERIS datasets should follow the following unified naming convention: [ANERIS\_dataset-name\_version\_creation-date], whereby data format should be DDMMYYYY, number style version should be 01, 02, 03.

*Example: ANERIS\_DatasetName\_v01\_06022023.xlsx*

Finally, to guarantee the registration, indexing, and discoverability of the project's (meta)data, ANERIS will upload it to reputable repositories equipped with search engines and indexing capabilities, such as [Zenodo](#). This ensures that the data can be harvested and made available as a searchable resource.

## 3.2 Making data openly accessible

ANERIS partners will strive to ensure unrestricted access to peer-reviewed scientific publications as well as all associated datasets that are directly linked to the project's outcomes and funded or co-funded by the project. It is mandatory for both the research paper and the underlying data to be made openly accessible as early as possible after the paper's publication, with the latest deadline being the conclusion of the reporting period in which the paper was published. In adherence to this requirement, partners will fulfill their obligation to deposit a machine-readable electronic copy of either the published version or the final manuscript accepted for publication in a recognized and trustworthy open access repository.



The ANERIS DMP has been developed by strictly following the provided by the European Commission DMP template for Horizon Europe projects, which is also accessible through the DMPOnline platform. While the platform allows for online development of the document, the ANERIS DMP has been developed in an offline document. DMPOnline also provides the option to openly publish the data management plan within the platform, hence, ANERIS will consider depositing the project's DMP to foster transparency and accessibility.

Additionally, the project Data Management Plan will be published in the specialised open access journal [Research Ideas and Outcomes](#) (RIO). Following the Open Science principles, the consortium will consider depositing its research outcomes, along with the DMP in a freely accessible collection in the RIO journal, which is launched and maintained by the project partner PENSOFT. The journal allows publishing of entire research cycles, starting with research ideas, grant proposals, data, software, methodologies, workflows, research and review papers, policy briefs and others (Mietchen et al. 2021). This will ensure that all project outputs are published in open access, with a stable DOI assigned and comprehensively collected in one place. The unique features of RIO Journal also allow for open and public post-publication peer review, encouraging discussion among peers and consecutively allowing the publication of updated versions of each article, linked via Cross Mark.

### Recommendations

ANERIS partners can choose to deposit their scientific publications in:

- trusted certified repository;
- trusted community-recognised repository;
- trusted institutional repository.

ANERIS members can choose between two ways of making their underlying data open access:

- They can upload the data to an open access research data repository under the latest available version of the Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain Dedication (CC 0) or a license with equivalent rights. They could use a generic repository such as Zenodo or a trusted institutional repository.
- They can publish datasets as open access data papers in an academic journal and then deposit them in an open access repository. For further information you can refer to Penev L, Mietchen D, Chavan V, Hagedorn G, Smith V, Shotton D, Ó Tuama É, Senderov V, Georgiev T, Stoev P, Groom Q, Remsen D, Edmunds S (2017) Strategies and guidelines for scholarly publishing of biodiversity data. *Research Ideas and Outcomes* 3: e12431. <https://doi.org/10.3897/rio.3.e12431>

### 3.3 Making data interoperable

In order to guarantee the effortless discoverability and seamless integration of datasets, ANERIS will adhere to a set of community-approved guidelines for interoperability. It will employ widely accepted metadata vocabularies, standards, formats, and ontologies that are commonly employed for representing knowledge. If there is a need for partners to utilise less common ontologies and vocabularies, mappings will be supplied to facilitate their alignment with more commonly employed ones. Moreover, partners will individually evaluate whether newly created ontologies or vocabularies (if any) should be openly shared, enabling others to reuse, enhance, or expand upon them. Overall, ANERIS will ensure data interoperability by utilising the services of OpenAIRE and/or the European Open Science Cloud (EOSC) metadata, as well as by incorporating interoperable mechanisms. The overall integration of the EOSC synergies will implement a pipeline integration and synergies between the project and EOSC. This includes the adoption and technical integration of relevant services from EOSC into the ANERIS architecture and aligning ANERIS results with the emerging EOSC interoperability guidelines. More details on the EOSC implementation can be found in D5.3 EOSC integration plan, which will be developed by EGI in M18 - June 2024.

The types of data format that the ANERIS partners are advised to use and follow are listed below:

The ANERIS consortium is recommended to utilise data formats that have derived from open source software, i.e. open file formats.

- data science and data analysis scripts, e.g. .R, .py
- structured text or mark-up file containing metadata information, e.g. DDI XML file, GeoRSS, Rmarkdown
- quantitative tabular data, e.g. comma-separated values (CSV) file (.csv) tab-delimited file (.tab)
- genomic data: .fasta, .fast5, .fastq
- image data: .jpg; jpeg, .RAW
- vector and raster data (essential: .shp, .shx, .dbf, optional: .prj, .sbx, .sbn); geo-referenced TIFF (.tif, .tiff, .tfw, .ddf) or CAD data (.dwg); tabular GIS attribute data

ANERIS will refrain from generating extensive databases that require access through Application Programming Interfaces (API). The project will ensure compatibility by employing well-established file formats listed in Table 1. By utilising these file formats, coupled with comprehensive, standardised, and machine-readable metadata (such as EML), and depositing them in trusted repositories with open access and programmatic capabilities (like Zenodo), ANERIS guarantees that its data can be interpreted by any programming language, without relying on proprietary software. Additionally, the project's data will include appropriate references to other relevant data sources, augmenting interoperability even further.

### 3.4 Increasing data re-use

The data reusability will be implemented with the use of open licences and open data mechanisms. Data generated within ANERIS will be licensed under the Open Data Commons Attribution License ODC-By. If authors wish to publish data related to a journal article under a different license than the Open Data Commons Attribution License (ODC-By), they must explicitly notify the Project Coordination. Other open data licenses that can be considered include Creative Commons CC0 (also referred to as "CC-Zero" or "CC-zero") and the Open Data Commons Public Domain Dedication and License (ODC-PDDL). Under the CC0 license, the person who associates a work with this license dedicates the work to the public domain, waiving all rights worldwide under copyright law, including related and neighboring rights, to the extent permitted by law. This means that anyone can copy, modify, distribute, and use the work, even for commercial purposes, without seeking permission. Publishing data under a non-attribution waiver like CC0 helps avoid potential issues of "attribution stacking" when aggregating data from multiple sources for reuse, especially if the reuse process is automated. In such cases, although there is no legal requirement to attribute the data creators, it is still considered best practice within academic citation norms to reference the data source, similar to referencing other research articles.

## 4. Other research outputs

In addition to ensuring the application of the FAIR data principles to the data generated and/or reused by ANERIS, the data management plan also encompasses other research outputs of the project. Recognising the significance of promoting openness and accessibility, the DMP outlines that, where possible, the listed outputs such as publications, datasets, software, and other relevant materials will adhere to the FAIR principles. By following these principles, the project aims to maximise the discoverability, accessibility, and potential for reuse of its research outputs, thereby fostering collaboration, transparency, and the advancement of knowledge within the research community and beyond.

The other research outputs of ANERIS encompass a wide range of marine-life related technologies, which will be developed within the context of WP2-WP5, acquisition systems, services, protocols, guidelines, software, algorithms, demonstration platforms and case studies, as well as AI services and frameworks.

## 5. Allocation of costs

Managing data in a FAIR manner is linked to various types of expenses. These costs can be classified into two primary categories: 1) article processing charges (APC) associated with publishing data in open access journals, and 2) fees for depositing data in global data repositories. While numerous data repositories allow authors to upload their data without any charges (such as Zenodo and GBIF), certain more general-purpose repositories like Dryad

impose fees on users for publishing their data, unless exemptions apply. ANERIS has diligently allocated costs among its partners, ensuring that each partner has a dedicated budget for open access publication expenses (refer to ANERIS Grant Agreement, ANNEX 2: ESTIMATED BUDGET FOR THE ACTION - direct costs). It is the responsibility of each ANERIS partner to spend this budget responsibly and prioritise open access publications.

## 6. Data security

The ANERIS consortium is prioritising the data security measures, henceforth, the partners have provided detailed information, concerning:

- data storage location
- server location
- back up procedures,
- data protection practices
- data protection officers in the institutions

### Recommendation

To ensure the responsible storage of personal data, the ANERIS members are advised to follow the below recommendations:

- inform respondents and participants of interviews, surveys and meetings about what will happen with their personal data and what are their rights in this respect;
- provide the respondents the opportunity to retrieve their response and/or have their entry deleted;
- include informed consent for data sharing and long-term preservation in questionnaires dealing with personal data;
- use personal information of respondents and participants of interviews, surveys and meetings only if they have given their consent for its use;
- anonymise/delete the personal information when it is no longer needed;
- provide access to personal information only on a need basis to fulfill tasks.

## 7. Intellectual property

The handling of intellectual property will adhere to the guidelines outlined in the ANERIS Grant Agreement. The Party that generates the results will have ownership over them. In cases where the results are jointly produced by two or more participants, the ownership will be shared among them. The terms of this shared ownership, as well as the protection and distribution of the results, will be determined through a written agreement on joint ownership. Each participant will be responsible for exploring potential avenues to protect results that may have commercial or industrial value. Further information on the IP management can be found in D6.3.

## 8. Ethics

The ANERIS consortium will actively address ethical concerns related to the involvement of human participants, particularly in the EU, including aspects such as informed consent, anonymity, and confidentiality. Going beyond mere formal compliance, ANERIS recognises ethics, including data ethics, as a crucial component of citizen engagement. This approach aims to provide explicit assurance to citizens that their privacy will not be compromised by their engagement with ANERIS, except in specific cases where individual consent will be sought explicitly and in advance. Additionally, ANERIS will ensure transparency by informing participants in advance about the data collection process, how it will be processed and utilised, and the intended purpose, while also reaffirming their right to revoke consent if circumstances change.

## References

European Commission, Regulation of the European Parliament and of the Council: Laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts, 2021.

<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>

Penev L, Mietchen D, Chavan V, Hagedorn G, Smith V, Shotton D, Ó Tuama É, Senderov V, Georgiev T, Stoev P, Groom Q, Remsen D, Edmunds S (2017) Strategies and guidelines for scholarly publishing of biodiversity data. *Research Ideas and Outcomes* 3: e12431.

<https://doi.org/10.3897/rio.3.e12431>

Mietchen D, Penev L, Georgiev T, Ovcharova B, Kostadinova I (2021) Open science in practice: 300 published research ideas and outcomes illustrate how RIO Journal facilitates engagement with the research process. *Research Ideas and Outcomes* 7: e68595.

<https://doi.org/10.3897/rio.7.e68595>

## Annex

📄 LimeSurvey - ANERIS Data Management Questionnaire.pdf