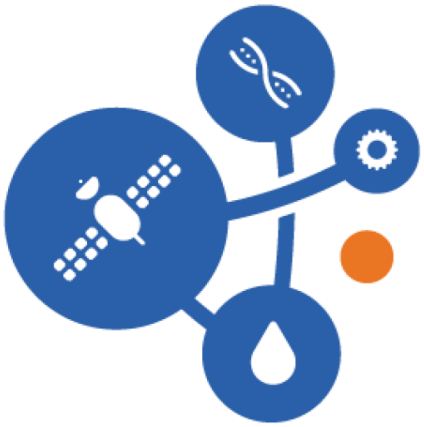

DiverSea

D 2.1 DATA MANAGEMENT PLAN
(preliminary)

Funded by
the European Union

DiverSea - HORIZON-CL6-2022-BIODIV-01-01
Grant Agreement 101082004

PROJECT	
Project Number	101082004
Project Acronym	DiverSea
Project Name	Integrated Observation, Mapping Monitoring and Prediction for Functional BioDiversity of Coastal SEAs
Call	HORIZON-CL6-2022-BIODIV-01
Topic	HORIZON-CL6-2022-BIODIV-01-01
Type of action	HORIZON-RIA
Service	REA/B/03
Project starting date	01 Sep 2023
Project duration	48 months

DELIVERABLE INFORMATION	
Number: D2.1	NAME: DATA NANAGEMENT PLAN
Responsible Partners	GATE-SU
Contributors	Denitsa Stoilova, Irena Pavlova, Glenn Dunshea
Type	DMP — Data Management Plan
Due Date	29/02/2025
Submission Date	01/03/2024
Dissemination	PU - Public

DOCUMENT HISTORY		
Revision Date	Modification	Author
23/02/2024	Initial Draft	Denitsa Stoilova, GATE-SU
26/02/2024	Review and Update	Irena Pavlova, GATE-SU
28/02/2024	Review and Update	Glenn Dunshea, NTNU
01/03/2024	Final version submitted	Irena Pavlova, GATE-SU

CLASSIFICATION	
Draft/Final	Final
Confidential/Restricted/ Public - Link	Public

TABLE OF CONTENTS

EXECUTIVE SUMMARY

DiverSea consortium is committed to performing professional management of data in the scope of the activities of the project. This document introduces the initial version of the Data Management Plan (DMP). It is deliverable D2.1 Data Management Plan and provides a general outline of the project policy for data management. The overall purpose of this DMP is to provide an analysis of the main elements of the data management policy that will be used by the project. The DMP describes the way in which the DiverSea consortium will manage the data that will emerge from the project, and how best practices in terms of metadata and archiving will be used to ensure that these data will be findable, accessible, interoperable, and reusable (FAIR) for other potential users. Furthermore, the DMP provides information about what data the consortium is aiming to preserve and in which format. The DMP is a living document that will be modified during the project lifetime with two more versions to be delivered later in the project

.

1 PROJECT INTRODUCTION AND OBJECTIVES

DiverSea, a Horizon Europe project, aims to revolutionize the field of marine observation and monitoring through the development of novel technologies and methodologies. By developing and integrating DNA-based identification approaches, such as "DNA-marks," with other advanced molecular techniques and ocean observation platforms, and AI-ML data processing architectures, DiverSea seeks to document essential biodiversity variables, monitor key biological parameters of marine species, and map critical marine habitats comprehensively.

The project's objectives are manifold:

1. **Innovate Marine Monitoring:** Harness low-coverage/cost genomic data to enhance species identification and understand genetic composition and species population essential biodiversity variables at a granular level.
2. **Quantify Uncertainties:** Develop methods to address uncertainties in marine eDNA/eRNA interpretation, leveraging new theories on molecular shedding and degradation.
3. **Integrate Techniques:** Combine emerging molecular techniques with autonomous systems, satellite remote sensing, and citizen science for a holistic view of marine ecosystems.
4. **Data Processing and AI:** Implement a novel AI-ML architecture for processing diverse biological, physical, and biogeochemical data, aiming for predictive insights into biodiversity and ecosystem dynamics.
5. **Stakeholder Engagement:** Develop an interactive platform, the "Biodiversity Services Dashboard," to communicate scientific outcomes to policymakers, stakeholders, and the public, visualizing key indicators and scenarios for biodiversity and ecosystem services.

DiverSea comprehensive approach to marine biodiversity monitoring and analysis promises to offer ground breaking insights into ecosystem functions, stressor impacts, and conservation strategies, directly informing policy and stakeholder decisions.

2 GLOSSARY, ACRONYMS AND ABBREVIATIONS

Short Name	Legal Name	Country
NTNU	NORGES TEKNISK-NATURVITENSKAPELIGE UNIVERSITET	NO
UCPH	KOBENHAVNS UNIVERSITET	DK
VITO	VLAAMSE INSTELLING VOOR TECHNOLOGISCH ONDERZOEK N.V.	BE
UNIVPM	UNIVERSITA POLITECNICA DELLE MARCHE	IT
LPN	Liga para a Protecção da Natureza	PT
UPF	UNIVERSIDAD POMPEU FABRA	ES
AT	ASOCIACION TURSIOPS	ES
DEU	DOKUZ EYLUL UNIVERSITY OF IZMIR	TR
IMB	INSTITUTE OF MARINE BIOLOGY OF THE NAS OF UKRAINE	UA
UPORTO	UNIVERSIDADE DO PORTO	PT
GATE-SU	SOFIA UNIVERSITY ST KLIMENT OHRIDSKI	BG
AUEB	ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS - RESEARCH CENTER	EL
UGOT	GOETEBORGS UNIVERSITET	SE
VNNOR	VNNOR AS	NO
NKUA	ETHNIKO KAI KAPODISTRIAKO PANEPISTIMIO ATHINON	EL
AKD	AKDENIZ KORUMA DERNEGI	TR
UG	UNIVERSITY OF GHANA	GH
INEGI	INEGI - INSTITUTO DE CIENCIA E INOVACAO EM ENGENHARIA MECANICA E ENGENHARIA INDUSTRIAL	PT
UCSD	The Regents of the Univ. of Calif., U.C. San Diego	US

Acronyms - Countries			
BE	Belgium	NO	Norway
BG	Bulgaria	PT	Portugal
DK	Denmark	SE	Sweden
EL	Greece	TR	Turkey
ES	Spain	UA	Ukraine
GH	Ghana	US	United States
IT	Italy		

DiverSea Project Terminology Abbreviation List

Abbreviation	**Meaning**
AI-ML	Artificial Intelligence & Machine Learning
API	Application Programming Interface
BHD	Birds and Habitats Directive
BSTA	Beyond State of The Art
CAPED	Create Assess Protect Exploit and Disseminate
CDOM	Chromophoric/Colored Dissolved Organic Matter
Copernicus	European Union's Earth observation program
CTD	Conductivity, Temperature, Depth
DNA & RNA	Deoxy Ribonucleic Acid & Ribonucleic Acid
DOP & DEP	Dissemination and Outreach Plan & Dissemination and Exploitation Plan
EC centre for biodiversity	European Commission Knowledge Centre for Biodiversity
eDNA/eRNA	Environmental DNA/RNA
EGS	Ecosystem Goods & Services
ELTER	Integrated European Long-Term Ecosystem, critical zone and socio-ecological Research
EMODnet	European Marine Observation and Data Network
EOV/EBVs	Essential Ocean and Biodiversity Variables
ESA	European Space Agency
GBIF	Global Biodiversity Information Facility
GEOBON	The Group on Earth Observations Biodiversity Observation Network
GEOSS	Group on Earth Observations
GOOS	Global Ocean Observing System
GUI	Graphical User Interface
ICES	International Council for the Exploration of the Sea
IoT	Internet of Things

IUCN	International Union for Conservation of Nature
KERs	Key Exploitable Results
KPI	Key Performance Indicators
LSTS	Laboratório de Sistemas e Tecnologia Subaquática
MARBIOSE	European MARine BIOdiversity and ecosystem goods & Services
MOU	Memorandum of Understanding
MPA	Marine Protected Area
MSFD	Marine Strategy Framework Directive
NCA	New Concepts and Approaches
NGO & SME	Non-Governmental Organization & Small and Medium-sized Enterprises
O/Obj.	Objectives
OBIS	Ocean Biodiversity Information System
OMoMaP	Observation, Monitoring, Mapping and Prediction
R&D	Research and Development
S & SS	Scope & Specific Scope
SEWP	Spreading Excellence and Widening Participation
SSH	Social Sciences and Humanities
STEM	Science, Technology, Engineering, and Mathematics
SWAFS	Science with and for Society
TRL	Technology Readiness Level
WFD	The Water Framework Directive
WP	Work Package

3 DELIVERABLE OVERVIEW

3.1 Document scope

The scope of this preliminary Data Management Plan is to describe the data produced in DiverSea and the methods and policies used for managing them. Within Task T2.1 Data Gathering**,** Data Management Plan (DMP) is elaborated and will be regularly updated, describing the management life cycle for the data to be collected, processed and/or generated. The plan follows EC Guidelines on FAIR Data Management and details the data the project will generate, how it will be exploited or made accessible for re-use, and how it will be preserved. DiverSea will also perform professional management of potential ethical issues related to project activities. Among others it will enable GDPR compliant data processing and alignment with other national and international regulations (see below). Furthermore, the plan will consider the EC's ethics guidelines for Big Data and regulations of EU High Level Expert Group on Ethical AI, as well as other ethics related aspects described in D7.1: The DiverSea Ethics Monitoring Plan. 3.2 Document status

The deliverable is of a public manner. From one side the deliverable is targeted for internal use within the consortium and is meant to support project partners to follow a unified and guided project implementation and data management process according to the guidelines and following the EC regulations and national law. On the other it is targeted towards the broader audience and mainly for DiverSea stakeholders, to provide for transparency in the data management process.

3.3 Document structure

The document starts with an executive summary, providing an overview of the document. The introduction and objectives of the project are discussed in Section 1, followed by the glossary, acronyms and abbreviations in Section 2. The purpose, scope, and context of the deliverable are set in Section 3. This section further delves into document scope, status, and structure. Section 4 explores the DiverSea Open Science Strategy, while DMP overall considerations are detailed in Section 5. Data-related aspects are covered in Section 6, which includes a data summary, data re-use, data types, data description, and purpose. The importance of FAIR data is highlighted in Section 7, where making data findable, accessible, interoperable, and increasing data re-use are discussed. Other research outputs are explained in section 8, while the allocation of resources and roles related to data management are addressed in Section 9. Section 10 focuses on data security, including provisions and long-term storage. Lastly, Section 11 highlights ethical and legal issues related to data sharing, informed consent, and long-term preservation and refers to D7.1 for other ethical aspects.

4 OPEN SCIENCE STRATEGY

DiverSea will actively contribute to the Open practices to provide access and stimulate the reuse of the research data generated by the project. Our methodology strongly relies on open collaboration with a large pool of external stakeholders (especially in WP4 and WP5), partners and relevant initiatives, such as COPERNICUS, EMODNET, Digital Twin of the ocean and others, representing European marine ecosystems – including academic institutions, research organizations, non-profit entities, public bodies and industry partners. This collaborative framework is designed to facilitate peer learning and best practice exchange, focusing on participatory activities in the form of consilience workshops, citizen science and crowdsourcing while aiming to reach all stakeholders of the targeted ecosystems. Particularly, the project seeks to engage young entrepreneurs, SMEs, technology suppliers, service providers, and consumers, as well as managers and policymakers involved in maritime and related sectors. The collaborative open science practices will be evident throughout the project lifecycle, from awareness-raising activities to the dissemination of results and their uptake by peers and stakeholders.

The project's activities, methodologies, and tools will be shared with the community from the onset through various communication and dissemination campaigns, as well as dedicated workshops to foster cooperation. This approach ensures continuous improvement through participation and feedback, enabling the direct involvement of external stakeholders across European marine and coastal ecosystems.

The results of our research will be made available through oral (events) and written channels (publications), in particular the organization of and the participation in community events. These measures will allow collecting feedback through open peer-review and consider the diversity of experiences for the continuous improvement of the methodology and its tools to benefit the community.

Moreover, all open data, results, publications, and documents produced by the DiverSea project will be made available through widely accessible platforms, including a dedicated project website hosted by NTNU. This website will offer free access to informational materials, including Fact Sheets and Visual/Audio content, and serve as a hub for disseminating policy-relevant syntheses of the project's outcomes. The website will also feature a connection to the project's innovative tools and dashboards, facilitating the integration of research findings into broader scientific, policy-making, and public engagement contexts.

Furthermore, all open data, results, publications and documents produced by the DiverSea project will be made available on the open access Zenodo repository (https://zenodo.org), where a dedicated community has been created for DiverSea: https://zenodo.org/records/10696929

More information on DiverSea Open Science strategies can be found in the sections below.

5 DMP OVERALL CONSIDERATIONS

One of the objectives of DiverSea is performing professional management of any issue that could emerge in the scope of the project activities. Therefore, this DMP has been elaborated, describing the data management procedures, rules, and policies for the data that will be collected, processed and/or generated during the project implementation by the project partners.

The plan also details what data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved. This plan has been prepared by taking into account the Horizon Europe for Data Management Plan template [v1.0 – 5 May 2021][1] , Guidelines on FAIR Data Management in Horizon 2020 [Version 3.0 of 26 July 2016][2] and the GDPR [Regulation (EU) 2016/679][3] . The processing of personal data will be carried out in compliance with the terms of the GDPR, under the legal basis of 'a task carried out in the public interest', and this will be documented as part of the ethics and management procedures of the project to be described in D7.1 and D6.2 accordingly. The project's reliance on this legal basis for the processing of personal data for research purposes will also be documented in its Privacy Notice. Personal data will only be processed where necessary to achieve the research and communication aims, and will be handled in accordance with data protection and research ethics principles.

Data subjects will be informed about the data processing and asked for their consent, when engaging users in workshops and participatory events – an information sheet will be provided for each case. In cases where existing datasets are analysed, the guidelines set forth by the Article 29 Working Party on purpose limitation will be considered. This means that data processing for the goal of statistical processing takes place separated from the original purposes for which this data was collected. The results of such processing are not linked back to the data subjects and all data is anonymized as early as possible.

Researchers will be trained in applying the necessary procedural safeguards and the project will take the necessary steps to ensure that applicable legislation concerning data protection is respected, and the steps taken during the project lifetime to ensure conformance to established guidelines will be documented.

DiverSea also includes an ethics requirements work package (WP7) to ensure compliance with all relevant ethics requirements.

The DMP is intended to be a living document in which information can be made available on a finer level of granularity through updates as the implementation of the project progress and when significant changes occur.

---

[1] Horizon-Europe-Data-Management-Plan-Template.pdf (enspire.science)

[2] h2020-hi-oa-data-mgt_en.pdf (europa.eu)

[3] REGULATION (EU) 2016/ 679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL - of 27 April 2016 - on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/ 46/ EC (General Data Protection Regulation) (europa.eu)

6 DATA SUMMARY

This DMP is developed in the scope of the DiverSea Data Space: Data integration, harmonization, cataloguing and compatibility with standards work package (WP2 – D2.1) and describes the collection and storage of the data, the free exploitation possibilities, for reproduction and dissemination purposes, by the consortium and third parties to access, as well as the tools made available to the beneficiaries. The data considered by DiverSea and this DMP are of various sources, type, and purposes. Along with the development of DiverSea project new data, not considered in this DMP version, may appear. This is why this DMP should be treated as a living document, will be revised, and updated accordingly.

The data management procedures for data acquisition – collection – management – processing – sharing are initially defined here, and will continuously be implemented and monitored during the entire project lifetime, and the publishing of the data will take place according to FAIR principles. Due to the nature of the project and the use of the outcomes by a wide audience of beneficiaries and stakeholders, some large part of the non-personal data generated during the project shall be made public unless there is some special reason to not allow for this e.g. in cases of confidential data (such as for example partners' exploitation plans). All results and progress on this Task will be reported and evaluated in this DMP.

6.1 Data re-use

The DiverSea project will extensively re-use existing data for a variety of purposes, enhancing its research and analysis capabilities. This re-use strategy includes integrating, harmonizing, and structuring a wide array of data sourced from both private and public datasets. These encompass historical and real-time data from project partners, alongside data derived from field studies and experiments, such as environmental DNA (eDNA), biological objects, and Essential Ocean Variables (EOVs) and Essential Biodiversity Variables (EBVs). To ensure the coherent integration and utilization of this data, the project will create a domain-specific vocabulary and data models that incorporate both new and existing vocabularies and ontologies. Examples include the Humboldt Core/Darwin Core, Sea Data Net, NERC Vocabulary Server, and Big Data Ocean. The objective behind this effort is to achieve data interoperability, allowing for the structured and standardized description of the collected data across the project.

Furthermore, DiverSea will implement data publication services to facilitate the availability of data within its Data Space, integrating different datasets and sources for subsequent modelling and analysis. This also involves ensuring the consistent quality of the data collected, aligned with the project's standards. Additionally, the project will offer data catalogue services, promoting data exploration and access while adhering to the FAIR data principles—findability, accessibility, interoperability, and reusability. Compliance with the GBIF guidelines for publishing eDNA-derived data is also a priority, aimed at enhancing the datasets semantic richness and accuracy.

DiverSea plans to integrate its Data Space with established EU data platforms and systems, such as EMODnet, Copernicus, and the European Space Agency Earth Observation Programme. This will enable the project to utilize a broad spectrum of essential data while ensuring compliance with established data standards, vocabularies, metadata formats, and data transport standards. The project's engagement with relevant projects like EUROPABON signifies its commitment to fostering data integration and maximizing the reusability of the data stored within its Data Space.

6.2 Data types

The DiverSea project will generate and re-use data in various types and formats to enrich its DiverSea Data Space. This encompasses a broad spectrum of data sources, including:

- **Private and Public Data:** Data from both private and public sources will be integrated into the DiverSea Data Space.
- **External Sources:** The project will incorporate data from external sources, enhancing the diversity and scope of its data repository.
- **Historical and Real-time Data:** It will utilize both historical data and real-time data from partners to provide comprehensive insights into marine biodiversity and ecosystems.
- **Field Studies of natural ecosystems and Experimental Data:** Specific examples include environmental DNA (eDNA), biological objects, Essential Ocean Variables (EOVs), and Essential Biodiversity Variables (EBVs).

In terms of data formats, the project will handle a wide variety of data types, such as:

- **Tabular Data and Relational Databases:** Structured data that is organized in tables or related collections.
- **Time-series Data:** Data that is collected over time, often from sensors and actuators.
- **Images:** Visual data from sources like satellites and drones.
- **Remote Sensing Data:** Data obtained from remote sensing technologies for environmental monitoring.
- **DNA sequence data** and summaries thereof: Data obtained from eDNA experiments and surveys and other molecular methodologies

To manage and process this diverse data, the project will employ several technologies and approaches:

- **Pre-processing of Data:** This includes filtering, curation, linking, imputation of missing values, standardization, handling inconsistencies, noise reduction, and anonymization where necessary.
- **Data Storage and Processing Infrastructure:** The project will use Hadoop Distributed File System and Cloudera software on a hybrid cloud to support data analytics and dashboard

functionalities. Additionally, NoSQL MongoDB and multimodal database ArangoDB will be used for storing various data types.

- **Data Conversion and Integration:** Adapters will be developed for converting raw sensor data from NoSQL databases into JSON format, facilitating the integration and analysis of the collected data.

The DiverSea will generate a diverse range of data types, which can be categorized into the following relevant categories based on the information provided and the nature of the project:

- **Raw Data:**
  - Environmental DNA (eDNA) samples and related biological objects.
  - Time-series data from sensors and actuators.
  - Images and remote sensing data from satellites and drones.
  - Species occurrence records.
- **Processed Data:**
  - Pre-processed eDNA data after filtering, curation, linking, imputation of missing values, standardization, inconsistencies handling, noise reduction, and, where necessary, anonymization.
  - Integrated datasets that combine historical and real-time data from partners with external sources, structured for analysis and visualization.
  - Harmonized data that has been standardized across different formats and sources to ensure interoperability and accessibility within the DiverSea Data Space.
- **Generated Data:**
  - New insights and knowledge derived from the analysis of raw and processed data, including predictive models of marine biodiversity and ecosystem health.
  - Data models and vocabularies developed to support data integration, harmonization, and structuring, including domain-specific ontologies.
- **Intellectual Assets:**
  - The methodologies, tools, and software developed for data processing, analysis, and visualization, including data integration and publication services.
  - Symbolic Knowledge Representation and Reasoning frameworks, including Knowledge Graphs built on Semantic Web standards.
- **Publications:**
  - Research papers, policy briefs, and reports generated from the project's findings, aimed at disseminating knowledge to the scientific community, policymakers, and the public.

The DiverSea project's focus on marine biodiversity and ecosystems, leveraging cutting-edge technologies and methodologies for data integration, ensures a comprehensive approach to data management. This categorization reflects the project's lifecycle from data collection and processing to the generation of new knowledge and intellectual assets, culminating in the dissemination of findings through publications.

6.3 Data description and purpose

6.3.1 Purpose of the data generation / re-use and its relation to the objectives of the project

The DiverSea project focuses on gathering and using both new and old data to deepen our understanding of the ocean's biodiversity. It creates new data, like eDNA analyses, to map out the variety of life under the sea, which is essential for protecting marine habitats and managing them sustainably. Re-using existing data complements this by providing historical insights, helping us track changes in marine life over time. This dual approach enriches our research base, offering a more detailed picture of marine ecosystems.

The data will come from different sources such as environmental DNA and RNA (eDNA & eRNA) samples, time-series data from sensors and actuators, satellite and drone imagery, species occurrence records from databases like GBIF, and genetic diversity information from the European Nucleotide Archive. These varied sources will enrich the DiverSea project's database, enabling comprehensive analyses of marine biodiversity and ecosystem health.

The data generated or re-used in the DiverSea project will be used for advancing scientific research on marine biodiversity, informing sustainable management practices for marine resources, enhancing conservation efforts, and fostering the development of innovative marine observation technologies. This comprehensive approach aims to contribute to a deeper understanding of marine ecosystems, support the protection of marine biodiversity, and ensure the sustainable use of marine resources, in line with global conservation and sustainability goals.

In essence, DiverSea aims to improve our understanding and conservation of marine environments through innovative data collection and sharing strategies, aligning closely with its core objectives of marine protection and sustainable management.

6.3.2 Expected size of generated/re-used data

Data size within DiverSea can vary widely depending on the type of data collected (e.g., high-resolution satellite images, genetic sequences, time-series data from sensors), the duration of data collection, and the scope of the project's research areas. Given the project's focus on marine biodiversity and ecosystems, which involves collecting and analyzing complex and diverse datasets, the overall data volume is likely to be substantial: DiverSea will handle large datasets typical of comprehensive environmental and biodiversity research efforts.

6.3.3 Origin/provenance of generated/re-used data

The origin or provenance of the data for the DiverSea project includes a mix of newly generated data and data re-used from various sources. Newly generated data come from activities such as collecting environmental DNA (eDNA) samples, conducting field studies and experiments, and utilizing advanced remote sensing technologies. This fresh data aims to provide new insights into marine biodiversity and ecosystems, contributing to the project's research goals. Alongside, the project re-uses existing datasets from a range of reputable sources. These include European dedicated research infrastructures like eLTER, which offers long-term biodiversity time series datasets from various EU marine environments, LifeWatch ERIC, and EMSO for comprehensive environmental monitoring data. The project also integrates data from global and European databases such as the Global Biodiversity Information Facility (GBIF) for species occurrence records and the European Nucleotide Archive for genetic diversity information. Additionally, specialized data platforms like EMODnet, Copernicus, and data from the European Space Agency Earth Observation Programme are utilized for their rich environmental and oceanographic data.

6.3.4 Data utility outside of the project

Table 1 illustrates what type of data will be collected, processed, or created in each specific Work Package (WP). It contains a description of the data that refers to each WP objective and explains the purpose of its collection. Additionally, the table indicates which data should be treated as open (public) and close (confidential).

Table 1 DiverSea data types, description and purpose in each Work Package

WORK PACKAGE	DATA TYPES	EXAMPLES	DESCRIPTION AND PURPOSE	OPEN / CLOSE
WP1: Developing nucleic acid methods, integrating field technologies and application to field case- studies	Raw data Processed data Generated data Intellectual assets and intellectual property Publications Personal data	• Field data collection. • Analytical documents: Evaluation of case studies • Project deliverables: reports on identification of case study-specific EBV/EOV and MSFD's variables, protocol of observation, monitoring, and sampling, field test-case study report, evaluation of case studies, report on eDNA/eRNA EBVs & indicators, and DNAmark development. • Technical information: development of DNAmark and methodologies for monitoring biodiversity. • Public reports on the findings and methodologies developed. • Stakeholders requirements	**Description:** WP1 task begins with molecular methods development, focusing on refining DNAmark techniques for genetic diversity mapping. This work involves creating reference libraries from genome skims of key species across Europe. These methods are then applied in various field case studies, integrating new data with ecological and biogeochemical processes to assess biodiversity Essential Biodiversity Variables. The data generated here feed into WP2 for integration into the DiverSea Data Space, which directly links data generation in WP1 with data management and application in subsequent WPs. **Purpose:** The data utilized in WP1 of the DiverSea project aims to develop DNAmark techniques for mapping genetic diversity and apply these to marine biodiversity variables. It also focuses on integrating advanced observation and monitoring systems across selected case study zones to enhance data acquisition and quality.	Open - all deliverables in WP1 are categorized as "Public" indicating that reports on case studies, protocols, evaluations, and developments in nucleic acid methods are fully open and accessible to the public. Open upon consent – personal data of the DiverSea stakeholders.
WP2: DiverSea Data Space: Data integration, harmonization, cataloguing and compatibility with standards	Raw data Processed data Generated data Public data dets Personal data	• Project deliverables: Data management plans (preliminary, updated, final), DiverSea Data Space (initial, final), including data lakes from existing portals and new data generated. • DSSE Data Space Blueprints: Initial and final versions of the DiverSea Data Space.	**Description:** WP2 focuses on the integration, harmonization, and cataloguing of biodiversity data, including that generated in WP1. T2.1, involves gathering diverse data types (e.g., eDNA, biological objects, EBVs) from multiple sources, including field studies from WP1. Raw sensor data will be converted from NoSQL databases into JSON… This task ensures that various data formats are pre-processed for quality and standardization, laying the groundwork for further analysis in WP3.	Open – data including the Data Management Plans (preliminary, updated, final) and the initial and final versions of the DiverSea Data Space. This means the data management strategies and data space developments are fully open to the public.

		• Personal data: gathered during Data Space evaluations by the stakeholders.	T 2.2 and T2.3 develop a domain vocabulary and metadata model to ensure interoperability across data sources, directly supporting data exploration in WP4's development of the MARBIOSE Dashboard. T 2.4 focuses on creating a data catalogue for easy data access, while T2.5 ensures the integration of DiverSea data with European and global data platforms. **Purpose:** To establish a unified, accessible data space that enhances the availability and usability of marine biodiversity data, fostering research, policy development, and public awareness.	Open upon consent – personal data of the DiverSea stakeholders.
WP3: Combining AI- ML and Statistical Methods for Predictive Scenario Building	Processed data Generated data Intellectual assets and intellectual property Publications	• Project deliverables: software tools for data acquisition and integration, ecosystem models, identification of key ecosystem properties, and development of indicators for the dashboard. • Technical information: details on software tools and algorithms used for data modeling and ecosystem analysis. • Analytical documents: reports on the development and application of software tools for ecosystem modeling.	**Description:** WP3 utilizes AI and ML to clean and model data from WP2, aiming to automate data cleaning and correct data collection biases. Task 3.1 develops tools for data quality checks and sampling process guidance, directly using data from WP2. This cleaned and structured data then serves as the foundation for Task 3.2, where ecosystem dynamics models are developed. These models are linked to ecosystem health indicators, which are crucial for WP4's dashboard development. Task 3.3 involves expert feedback to refine these indicators, ensuring they accurately reflect ecosystem health for use in WP4's MARBIOSE Dashboard. **Purpose:** To develop sophisticated tools and models that predict marine ecosystem dynamics, providing essential insights for biodiversity management and conservation efforts.	Open - all data in WP3, which include software tools for data modelling and integration as well as ecosystem models. This will ensure open access to these tools and models developed for predictive scenario building.
WP4: Knowledge Integration: The MARBIOSE Dashboard for Biodiversity and	Processed data Generated data Intellectual assets and intellectual property Personal data	• Project deliverables: Holistic system model linking coastal and marine EGS to biodiversity, Dashboard tools as proof-of-concept, marine biodiversity, and ecosystem goods and services reports, operational dashboard,	**Description:** WP4 integrates data and insights from previous WPs to develop the MARBIOSE Dashboard, an interactive tool for exploring the relationships between marine biodiversity, ecosystem services, and human well-being. It starts with stakeholder mapping in Task 4.1 to identify key ecosystem services and drivers of change (Task 4.2),	Open – all results of WP4 will be open. These are the holistic system model, dashboard tools, and reports on marine biodiversity and ecosystem services, as well as the final report on the MARBIOSE dashboard.

Ecosystem Goods and Services	Publications	MARBIOSE dashboard – Final report. • Requirements information - Stakeholder Mapping and analysis to design of mental models for the causal dependencies between the drivers of marine biodiversity and related EGS • Technical information: Reports on MARBIOSE dashboard, including system models, design, functionalities, and application range. • Analytical documents: Final report on the MARBIOSE dashboard, detailing the integration of data and models for visualizing biodiversity and ecosystem service	which are informed by data and models from WPs 1-3. The systemic knowledge base developed in Task 4.3 utilizes data from WP2 and predictive models from WP3, while Task 4.4 focuses on the technical and functional design of the dashboard. **Purpose:** To provide a powerful tool for policymakers, scientists, and the public to explore and understand the complex relationships between marine biodiversity, ecosystem services, and human well-being.	Open upon consent – personal data of the DiverSea stakeholders.
WP5: Consilience, Public Interface, Communication, Dissemination & Exploitation	Generated data Intellectual assets and intellectual property Personal data Publications	• Project deliverables: consilience ontology protocol and evaluation/actions reports, Dissemination Exploitation & Outreach Plan with revisions, policy-relevant syntheses & policy briefs, reports on citizen science initiatives, SME knowledge transfer report. • Conference, workshop presentations, training materials • Social media • Project marketing materials • Visual/Audio Material • Logo and branding	**Description:** WP5 is dedicated to disseminating and exploiting the project's findings, engaging with the public, and promoting biodiversity conservation. It integrates outcomes from all other WPs, particularly the tools and data from WP1-4, to enhance public understanding and engagement in marine biodiversity issues. Tasks include developing a comprehensive communication strategy, maintaining a project website, and organizing citizen science initiatives to foster ocean literacy. **Purpose:** The data collected through WP5 will be used to ensure that the project's scientific achievements are effectively communicated to a broad audience, including policymakers, stakeholders, and the general public, maximizing the impact of DiverSea research and tools on marine conservation efforts.	Open – all communication and dissemination material Open upon consent – personal data of the DiverSea events participants

WP6: Project Coordination	Generated data Personal data of partner personnel	• Project deliverables • Meeting notes/minutes • Administrative documentation	**Description:** It will be used to facilitate efficient implementation, optimize resources, monitor tasks, ensure legal compliance, and produce reports for the European Commission. **Purpose:** The data will be used to facilitate effective project management, ensuring compliance with legal and ethical standards, and supporting decision-making processes within the project's governance structure. It will also be used to produce reports for the European Commission.	Closed – confidential data related to internal project management (e.g. project deliverables flagged as sensitive: **D6.1** - Templates for task planning and reporting on scientific and financial issues, **D6.2** - Guidelines for report and deliverables collation. Closed data is also personal data.
WP7: Ethics requirements	Personal data	• Project deliverables: Ethics monitoring plan, ethics deliverables, and ethics checks. • Meeting notes/minutes • Administrative documentation	**Description:** In WP7 the emphasis is on ethics requirements and monitoring to ensure all activities comply with ethical standards. It involves creating and implementing detailed ethics monitoring plans, conducting regular checks to ensure adherence to these plans. These efforts are aimed at addressing and maintaining ethical standards across all project activities, ensuring the protection of participants, the integrity of data, and compliance with legal norms. **Purpose:** to uphold and monitor the ethical standards of the project, ensuring all research activities involving data adhere to established ethical guidelines and contribute to the project's integrity.	Closed – confidential data related to ethics requirements (e.g. project deliverables flagged as sensitive), personal data filled in the informed consent forms

7 FAIR DATA

The DMP aligns with the "FAIR" principles, meaning that the projects' data will be made findable, accessible, interoperable, and re-usable. To incorporate "FAIR" principles, the Data Management Plan considers the following:

- what data will be generated and/or generated, and processed,
- the handling of data,
- which methodologies and standards will be applied and implemented,
- which data will be shared or made open access and how,
- how data will be curated and preserved.

DiverSea endorses the EC's motto: "to make the data as open as possible, but as closed as necessary". As presented in the table in the previous section, the privacy of the participants involved, and the confidentiality of specific results or valuable operational and business knowledge will be protected. In these cases, the data collected, processed, or produced by the project will not be made available for public use. In all other cases, the data will be made available as broadly as possible.

7.1 Making data findable

All data and documents produced by the DiverSea project and flagged as confidential will be only findable by consortium members and EC in internal management and storage system (MS Teams). Files stored in the system should be described by the following naming structure:

filename_version_uploaddate

and project deliverables:

deliverablenumber_deliverableshortname_version_uploaddate.

All public documents and project deliverables will be put on the project website and made easily findable in its structure.

As mentioned, all open data and documents produced by the DiverSea project will be made available on the open access Zenodo repository (https://zenodo.org) and will be identified by a persistent identifier (PID). The PID used will be the Digital Object Identifier (DOI). DOI is a unique alphanumeric string assigned by a registration agency (the International DOI Foundation) to identify content and provide a persistent link to its location on the Internet. The publisher assigns a DOI when an article or other object is published and made available electronically.

Rich metadata – creation and standards

Rich metadata will be provided to allow easy discovery of DiverSea data produced. The metadata that will be created will include for example title, authors and their organizations, publication

date, publication venue, abstract (if applicable), keywords, license, version, access rights, related resources (such as links to related publications, datasets, software, or others), references, Horizon Europe funding acknowledgement, project name, acronym and grant number, persistent identifiers. In Zenodo, that has been selected as an external open repository (see above), the project partners will be able also to deposit files, images, documentation, datasets, or source code. These can also have their own metadata associated with them, providing further context and description. Regarding disciplinary or general standards, the metadata will be created using standard metadata schemas such as

DiverSea results and public outcomes are also visible in the EC data basis of research results CORDIS: https://cordis.europa.eu/project/id/101082004 .

Search keywords will be provided in the metadata to optimize the possibility for discovery and potential re-use of the data collected. Metadata will be offered in a way that can be harvested and indexed, as this is a common practice for research data management.

7.2 Making data accessible

All data related to the project and exchanged within the consortium, including intermediate versions of the deliverables, meetings' material (agenda, notes, presentations, demos, minutes, etc.) and any other documents used for gathering inputs from the project's partners will be uploaded and stored on the internal data management and storage system - MS Teams.

As for the publications, where the analyses of the empirical research data coming out of the DiverSea project, will be presented, the consortium will publish them in scientific journals. Journals that allow open access will be preferably chosen for publishing activities. Therefore, the publications will be stored both in the project data management system and publisher repository.

All the other public data produced by the project will be made available through the project website, EC platforms such as CORDIS and Zenodo open repository.

7.2.1 Usage of a trusted repository

The partners of the DiverSea project will ensure open access to peer-reviewed scientific publications related to their results. This includes depositing a machine-readable electronic copy of the published version or the final peer-reviewed manuscript accepted for publication in a trusted repository for scientific publications at the latest time of publication. Immediate open access to the deposited publication via the repository will be provided under the latest available version of the Creative Commons Attribution International Public License (CC BY) or a license with equivalent rights. For monographs and other long-text formats, the license may exclude commercial uses and derivative works (e.g., CC BY-NC, CC BY-ND). The repository will also provide information about any research output, or any other tools and instruments needed to validate the conclusions of the scientific publication.

The project consortium has decided to use the multidisciplinary repository Zenodo (https://zenodo.org). Zenodo is a popular open-access repository that allows researchers to share and preserve their scientific outputs and data. It is the recommended "catch-all" repository for projects like DiverSea that do not have immediate access to an organized data centre, according to OpenAIRE's (https://www.openaire.eu) recommendations. The DiverSea community has been created: https://zenodo.org/records/10696929 .

7.2.2 Assignment of a data identifier including an identifier to a digital object

DiverSea will offer Open Access (OA) to its results reported in publications, as partners consider that by giving OA to scientific results and publications, important breakthroughs can be speeded up in the EU industries and companies, boosting knowledge and competitiveness. All the publications will have a Digital Object Identifier (DOI).

Moreover, Zenodo assigns persistent identifiers to the deposited content to ensure long-term accessibility and citability. The primary PID used by Zenodo is the DOI.

7.2.3 Open access to data and restricted access conditions

The DiverSea project is committed to open practices and aims to provide access to and encourage the reuse of research data generated by the project. The project methodology emphasizes open collaboration with a diverse range of external partners, including academic institutions, research organizations, non-profit entities, SMEs, industry, policy makers, and public authorities. The project will focus on engaging these stakeholders within the targeted ecosystems, particularly organizations that can support the project's development. By applying collaborative open science practices throughout its life cycle, from awareness-raising to dissemination and the uptake of results by peers, DiverSea seeks to involve all relevant stakeholders. This approach aims to foster a broad engagement, leveraging the support and expertise of these partners to ensure the project's outcomes are widely adopted and contribute to the conservation and sustainable management of marine biodiversity.

While the project is committed to open science practices, there may be circumstances where some data cannot be made openly available due to legal, ethical, or business reasons. In such cases, the project may need to impose restrictions on access to the data or apply embargoes for a limited period of time. As an example, due to legal and business reasons the exploitation plans of the partners will be kept confidential.

DiverSea project will make a significant amount of research data openly available in accordance with best practices for open science and collaboration.

Based on the DiverSea project's data, it may be necessary to apply an embargo to some research data to allow sufficient time for preparing manuscripts for publication or for filling patent applications. The specific duration of any embargo will be determined based on the specific needs of the research and will be clearly justified in writing by the researcher or research team. However, any re-used data will be properly cited and attributed in accordance with best practices

for data management. It is important to note that during an embargo period, metadata describing the research data will still be made available as soon as possible to ensure transparency and enable discovery of the research.

7.2.4 Access to data in cases of restrictions on use

If it is necessary to restrict open access to research or other data, efforts will be made to provide controlled access to these data. In cases where restricted or embargoed data is stored in the Zenodo repository, information about the restricted data will be published in the repository, and details about when the data will become available will be included in the metadata. Metadata for both open and closed records are always publicly available in Zenodo. However, data files and datasets for restricted access records are only visible to their owners and those who have been granted access by the owner. Restricted access allows to upload data and provide conditions under which they grant access to these data. External users who wish to request access must provide a justification for how they fulfil these conditions to the Zenodo community manager. The owner of the data is also notified of each new request and can decide whether to accept or reject it. If the request is accepted, the requester receives a secret link that usually expires within 1-12 months.

If there are restrictions on the use of data in the DiverSea project, access to the data will be provided under controlled conditions to individual researchers. During the project, access to restricted data will be granted to those who have been authorized by the owner of the data.

After the end of the DiverSea project, access to restricted data will continue to be provided under controlled conditions. Information about the restricted data will be published in Zenodo repository, and details about when the data will become available will be included in the metadata. Access records are only visible to their owners and those who have been granted access by the owner. To request access to restricted data, researchers must provide a justification for how they fulfil the conditions set by the data owner. The owner of the data is notified of each new request and can decide whether to accept or reject it. If the request is accepted, the requester receives a secret link that typically expires within 1-12 months.

All project data will be also available through the internal collaboration platform – MS Teams. Confidential and personal data will be shared only with authorized consortium partner members through secure channels. Data meant for the open public will be part of the content published on the project website. The data sharing logic adheres to the FAIR principles and follow the arrangement described below:

- Project deliverables marked as sensitive in the Description of Action and meeting minutes are confidential and will be accessible only through the internal collaboration platform – MS Teams only by the consortium members (including the Commission Services).
- Project deliverables marked as public will be accessible on the project public web site, on Zenodo and on the EC CORDIS system.

- Conference and workshop presentations, as well as surveys and interview notes, will be uploaded on the project public web site, unless they are flagged as confidential and then made available only through the internal document management system.
- Dissemination material (press releases, leaflets, logo, etc.) will be distributed or presented in internal and external events and will be made available via the project public website and social media accounts and the consortium partners official communication channels.
- The identity of the person accessing the data will be ascertained through:
    - User Authentication: Users will be required to provide valid credentials, such as a username and password, to access the system. This will verify the identity of the person accessing the data.
    - Access Control Lists (ACLs): Access control lists will be used to define and manage the permissions granted to specific individuals or groups. By assigning specific roles and permissions to users, the project will restrict access to data based on the identity of the person attempting to access it.

7.2.5 Metadata and its availability

Metadata of deposited data and publications will be open under a Creative Common Public Domain Dedication (CC 0) or equivalent, in line with the FAIR principles. Metadata will contain information to enable users to access the data. This will include descriptive information such as title, author, date, location, and other relevant information as stated above that can help users identify and locate the data they need. Additionally, metadata may include information about access permissions and user restrictions, which can help ensure that the data is only accessed by authorized users.

The project consortium will make sure metadata remains available even after the data is no longer available. This will be achieved by depositing metadata in the trusted Zenodo repository, where it can be preserved and curated for long-term access and re-use. Metadata will be stored in a machine-readable format to ensure that it can be easily accessed and used by future researchers.

7.3 Making data interoperable

7.3.1 Data and metadata vocabularies, standards, formats or methodologies

The DiverSea project aims to ensure the interoperability of its data by adhering to community-endorsed best practices and standards. This includes employing widely recognized data and metadata vocabularies, standards, formats, and methodologies that facilitate data exchange and reuse within and across disciplines. Specifically, the project will follow established standards such as Darwin Core for biodiversity data, the Ecological Metadata Language (EML), and others relevant to marine and biodiversity research.

7.3.2 Usage of mappings to ontologies

In the DiverSea project, recognizing the importance of interoperability and the reuse of data across different scientific domains, we will employ mappings to established ontologies to ensure our data is compatible with community-endorsed standards. This approach facilitates the exchange and integration of our project's data with broader datasets, enhancing the potential for multidisciplinary research collaborations. When specific or novel data types necessitate the development of project-specific ontologies or vocabularies, we commit to providing mappings to more widely used ontologies. This ensures that our data remains accessible and usable within the broader scientific community. Furthermore, the DiverSea project will openly publish any newly generated ontologies or vocabularies, encouraging their reuse, refinement, or extension by other researchers.

7.3.3 Qualified references to other data

Based on the FAIR data principles, data will include qualified references to other data to optimize data reuse and allow for replication and/or combination of data in different settings. This means that metadata and data will be well-described, and that metadata will include references to other relevant data sources, including datasets from previous research or other data from the DiverSea project itself. This strategy not only enhances the project's data ecosystem but also fosters a collaborative environment conducive to advancing marine biodiversity research.

7.4 Increase data re-use

The data produced by the DiverSea project will also be useful for other related policies, projects and initiatives both at national and EU levels.

7.4.1 Documentation needed for validation of data analysis and facilitation of data re-use

In the DiverSea project, documentation essential for validating data analysis and facilitating data re-use will be meticulously prepared and provided. This documentation will include comprehensive readme files that detail the methodology employed in data collection and analysis, codebooks for understanding the data structure and variables, information on data cleaning procedures, and analyses undertaken. Additionally, it will cover variable definitions and units of measurement to ensure clarity and consistency across datasets. This approach is designed to make the project’s data not only accessible but also understandable and reusable for future researchers, supporting the principles of open science and data stewardship. By ensuring that all necessary documentation is in place, DiverSea aims to enhance the transparency of its research processes and enable effective data re-use within the scientific community, contributing to the advancement of marine biodiversity research and conservation efforts.

Usability of data by third parties after the end of the project

Open results produced by the project and deposited in the respective Zenodo repository will be usable by third parties after the end of the project. If confidentiality, security, personal data protection obligations or IPR issues related to specific research data that is needed to validate a scientific publication forbid open access, the data may be deposited in a restricted repository and access may be granted upon request and under the conditions of a restricted license.

7.4.2 Documentation of data provenance

Data provenance is an important aspect of the DiverSea project's data management plan. Provenance documents the inputs, entities, systems, and processes that influence data of interest, providing a historical record of the data and its origins.

To document data provenance, the DiverSea project team will include a Provenance Description in their data documentation and in the respective metadata. This description will detail the process that led to the production of the current data, including input data, links to datasets (if possible), and a description of the process or model used to process the input data and produce the data. If applicable, the description will also include a link to the model used to produce the data.

The project team will also include links to any external datasets used as a basis for the creation of the current data in their documentation. Additionally, if possible, they will include links to any internal data that were used to produce the current data. If the current data is produced by a model, these links would point to the "model input" data.

7.4.3 Data quality assurance processes

For the DiverSea project, ensuring the highest data quality is paramount for the credibility and reliability of its research outcomes. The project will implement a comprehensive data quality assurance process, encompassing several key strategies:

1. **Standardization of Data Collection Methods:** DiverSea will, where possible without compromising the integrity of time series data in field studies, utilize standardized protocols for data collection across all project activities, ensuring consistency and comparability of data. This includes the use of validated methodologies and tools for environmental sampling, data recording, and analysis.
2. **Use of Trusted Data Sources:** When re-using existing data, the project will prioritize data from reputable, peer-reviewed sources. This ensures that the data integrated into the project's databases meets established quality standards.
3. **Rigorous Data Validation and Verification:** Data collected and processed by the project will undergo thorough validation and verification processes. This includes checks for accuracy, completeness, and consistency, using automated tools where possible to identify and correct errors or anomalies.

4. **Peer Review of Data:** Key datasets generated by the project will be subject to peer review by experts within and outside the consortium. This external validation helps to identify any potential issues or biases in the data collection and analysis processes.
5. **Metadata Documentation:** Comprehensive metadata documentation will accompany all datasets, detailing the methodology used, data collection parameters, analysis techniques, and any assumptions made. This ensures transparency and facilitates the replication and validation of research findings by others.
6. **Training and Capacity Building:** Project team members will receive training in best practices for data quality assurance, including the use of specific tools and techniques for data validation. This builds internal capacity to maintain high data quality standards throughout the project.
7. **Continuous Monitoring and Improvement:** The project will implement ongoing monitoring of data quality, with mechanisms in place to address any issues identified promptly. Feedback from data users will also be solicited to inform continuous improvement efforts.

8 OTHER RESEARCH OUTPUTS

In addition to the management of data, DiverSea team also considers and plans for the management of other research outputs that will be generated or re-used throughout or after the project. The different research outputs produced by the projects are detailed in Table 1. The access to them will be allowed using FAIR data principle so that they be used in diverse context.

9 RESOURCES AND RESPONSIBILITIES

9.1 Allocation of resources

The scientific publications, where the analyses of the DiverSea research data will be presented, will be published in scientific journals that allow open access. The costs related to open access will be claimed as part of the Horizon Europe grant.

9.2 Assessment of costs needed to make data or other research outputs FAIR

There are no immediate costs anticipated to make the open results generated in DiverSea FAIR. Especially no costs are foreseen for storing open results in the project's default repository (Zenodo). Additional details will be reported, as needed, in future versions of the DMP.

Any unforeseen costs related to open access to research data in Horizon Europe are eligible for reimbursement during the duration of the project under the conditions defined in the Grant Agreement.

9.3 Roles related to data management

Data management activities concern the whole project and need to be coordinated and monitored both at project and work package level. Data management is also linked to publication of project results and thus dissemination activities. Therefore, the following roles and responsibilities can be identified:

The project coordinator, project management team, and WP7 leader are jointly responsible for verifying the implementation of data protection measures throughout the project. They review consent forms, information leaflets, and other documents to ensure compliance with GDPR. They provide training and guidance to all project members who handle personal data, particularly the interviewers. They act as the primary contact for individuals with inquiries regarding privacy and data protection. Additionally, they ensure prompt deletion of data if consent is revoked at any point.

GATE-SU is responsible for the development of the Data Management Plan.

The WP and Task Leaders are responsible for:

- the implementation of the data management policy in their respective WPs,
- monitoring data management activities and deadlines and sending reminders to partners,
- asking partners for missing information or clarifications,
- providing input to the data management plan by analysing and summarising the WP-specific data surveys,
- offering customized help and further guidance for publishing open data and for filling out the WP data surveys,
- monitoring that open results (data and software) are deposited in the default internal repository and the complementary OpenAIRE-compliant Zenodo repository and sending reminders to partners,
- monitoring that open results available in OpenAIRE are properly linked with DiverSea (see https://www.OpenAIRE.eu/participate/claim).

10 SECURITY

10.1 Data security provisions

The Zenodo repository houses open results, which are securely stored in the CERN Data Centre. Multiple replicas, both for data files and metadata, are maintained independently online. CERN possesses extensive expertise in constructing and managing expansive digital repositories and is dedicated to upholding the data centre to accommodate the increasing LHC data, which is expected to grow to hundreds of petabytes over the next two decades. The servers are managed according to the CERN Security Baseline for Servers. For more information see http://about.zenodo.org/infrastructure/.

The DiverSea project involves the collection and processing of personal data, defined as any information relating to an identified or identifiable natural person, such as names, addresses, and email addresses. All partners in the project are required to follow ethical guidelines / requirements as described in D7.1, ensuring that personal data is not disclosed without informed consent from the individuals involved. If it required to be at all, personal data will only be made publicly available after obtaining explicit informed consent from the stakeholders and implementing necessary measures such as aggregation and anonymization to protect individuals' privacy.

To safeguard data security, including recovery, secure storage/archiving, and the transfer of sensitive data, the project will implement a comprehensive Ethical monitoring plan (see D7.1). These guidelines will detail the procedures for ethical compliance, provide templates for consent forms in multiple languages, and outline measures for secure data handling. This framework ensures that all aspects of personal data management, from collection to eventual archiving or publication, are conducted with the utmost care for privacy, security, and ethical considerations, reflecting the project's commitment to responsible data management practices.

10. 2 Data storage for long term preservation and curation

Apart from the internal project data management system, the trusted repository used in DiverSea will be Zenodo. Zenodo is considered a trusted repository for long-term preservation and curation of research data. Zenodo is a general-purpose open repository developed under the European OpenAIRE program and operated by CERN. It allows researchers to deposit research papers, data sets, research software, reports, and any other research-related digital artefacts. For each submission, a persistent digital object identifier (DOI) is minted, which makes the deposit citable and trackable. Zenodo is built to ensure that everyone can join in Open Science. The repository assumes responsibility for the long-term preservation and manages this function in a planned and documented way and has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end-users to make quality-related evaluations.

11 ETHICS

11.1 Ethical or legal issues to impact data sharing

The DiverSea project includes comprehensive measures to address ethics and legal issues, particularly in relation to data sharing and long-term preservation. Specifically, WP7 is dedicated to ensuring compliance with ethics requirements throughout the project's duration. This includes adherence to legal and regulatory standards, safeguarding personal data, and ensuring informed consent for data sharing and preservation, especially in questionnaires dealing with personal data. All currently identified ethical parameters to be adhered to will be documented in Deliverable D7.1 DiverSea Ethics monitoring plan, Deliverable D7.2 EPQ - NEC - POPD - H - Requirement No. 2 and Deliverable D7.3 EPQ - POPD - NEC - H - Requirement No. 3.

DiverSea applies the following strategies for data collection, to address the data protection issues involved with handling of personal data:

- The stakeholders involved in DiverSea activities are exhaustively informed so that they can autonomously decide whether they consent to participate or not (see Annexes). The purposes of DiverSea, the procedures, as well as the handling of their data, is explained (See Appendix 3, D7.1).
- The data exploitation is in line with the respective national and EU data protection acts. DiverSea team will anonymize all personal data to avoid tracing data back to individuals that might become this way identifiable and the data to be abused. In case participants do not consent to reveal their personal data, it will be stored in anonymous forms, so the identities of the participants will only be known by the project partners involved. Raw data that may reveal the identity of individual participants will not be shared with third parties. These privacy guidelines will ensure that personal data is treated with the utmost care to ensure the privacy of individuals.
- The data will be used only within the project framework of DiverSea and will not be made accessible for any third party.
- All files stored on the collaboration platform will only be accessible to the consortium partners and protected by an authentication mechanism. These files will be stored on a private server and backed up regularly. Unauthorized users will have no read, edit or write access to these files, nor to the files structure.

11.2 Informed consent for data sharing and long-term preservation

Deliverable D7.1 NEC - EPQ - POPD - H - Requirement No. 1, Deliverable D7.2 EPQ - NEC - POPD - H - Requirement No. 2 and Deliverable D7.3 EPQ - POPD - NEC - H - Requirement No. 3 provide information about applying ethical standards and requirements. They also inform DiverSea project partners when and where approvals are needed and how to deal with personal data. Further on, it is communicated where informed consent is needed and how to get them, e.g. addressing informed consent procedure for communication with stakeholders.

Participants participatory activities will be provided with a consent form and information sheet (see Annexes) that outlines the purpose of the research, how their personal data will be used, and how long it will be preserved.

ANNEX I CONFIDENTIALITY AGREEMENT

DiverSea Confidentiality Agreement

The stakeholders' data shared between the partners in the DiverSea project contains personal information, and therefore its usage is protected by law. To comply with this law, usage and sharing is restricted and you should follow the rules and guidelines described in the ethical guidelines and data management plan defined for DiverSea for collecting, processing, sharing and storage of data.

In addition to this you have to accept the following terms:

- I will not share the data I received with any third parties, including the employers of the participants, or other members of the consortium of the DiverSea project without explicit consent from whom I received the data.
- I will instruct the people listed who have access to the data and ensure that they follow the guidelines defined for the project.
- I will delete the data at the latest _____months after the results have first been published. (recommended time is 3 months).

Declaration on consent:

I hereby declare my consent with the rules described above:

Date: .....

Name & Organization: .....

Signature: .....

List of persons in my organization who have access to the data:

Person name	**Role in the project**

ANNEX II DIVERSEA DATA EXCHANGE FORM

DiverSea Data Exchange Form

As partners in the DiverSea consortium we agree to share personal information of individuals and/or organisations as defined in the Confidentiality Agreement.

This data exchange form documents the exchange between partners regarding personal data.

Team member(s) responsible for the data (who collected the data originally)	
Type of data (e.g. interview recording, questionnaires, workshops, etc.)	
Consent form signed by all involved participants	
Storage location	
Person(s) who have access to the data	
Purpose of sharing	
Confidentiality agreement signed	
Data retention (timeframe for storing the shared data)	

Date: .....

Name & Organization: .....

Signature: .....

ANNEX III DIVERSEA INFORMED CONSENT

DiverSea Declaration of Consent

The data collection is part of the research and innovation activities of DiverSea - Integrated Observation, Mapping, Monitoring and Prediction for Functional BioDiversity of Coastal SEAs (Grant agreement No. 101082004) funded under Horizon Europe program.

Your data will be held and used on an anonymous basis only for the purpose of the project DiverSea. It will not be stored after the end of the project unless required by specific national legislation. Your raw data will be kept confidentially and not disclosed to third parties. Compiled reports based on the collected information won't contain any personal data or data that could lead to the identification of a specific data subject.

The participation is voluntary, consent can be refused, and withdrawal is possible at any time.

Declaration of consent: I hereby declare my consent for my data to be conveyed and documented for the above stated purpose. I confirm that my participation is voluntary. I am aware that I may withdraw my consent at any time.

Date: .....

Name & Organisation: .....

Signature: .....

Signature DiverSea representative: .....

For further information about DiverSea project, please contact contact@gate-ai.eu

ANNEX IV DIVERSEA INFORMATION SHEET TEMPLATE

DiverSea Information Sheet

1. DiverSea Research Project Title

The title will be self-explanatory and will refer to concrete DiverSea internal, flagship, contract or collaborative project.

2. Invitation Paragraph

To explain why the potential participant is being asked to take part in a research project and its concrete activity such as interview, focus group, workshop, round table, or other event.

Example paragraph (all example paragraphs below are taken from UCL Research Ethics Committee[4]):

'You are being invited to take part in a research and innovation action project. Before you decided it is important for you to understand why the activities in DiverSea are being done and what participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.'

3. Project objectives and scope

Short summary of the project will be provided to put the participant in context. This will include the background, objectives, activities and duration of the project.

4. Why the participant has been chosen?

The detail of the inclusion criteria will be provided, together with an explanation of how the participant was chosen and how many other participants will be involved in the activity and the project in general.

[4] https://ethics.grad.ucl.ac.uk/forms/model-participant-information-sheet-and-guidance.pdf

5. Does the participant has to take part?

It will be outlined that taking part in the activity is entirely voluntary and that refusal to agree to participate will involve no penalty or loss of benefits to which the participant is otherwise entitled. The participant may discontinue participation at any time without penalty or loss of benefits to which the participant is otherwise entitled.

Example paragraph:

'It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. You can withdraw at any time without giving a reason and without it affecting any benefits that you are entitled to.'

6. What will happen to the participant if taking part in the activity?

It will be explained for how long the participant will be involved in the activity or research, how long the activity will last (if this is different), how often they will need to participate and for how long each time. Travel expenses will be reimbursed arrangement will be also specified.

The specifics of the activity and how it will be performed will be also detailed.

7. What is the concrete role of the participant?

If there are any lifestyle restrictions as a result of participating they will be described here.

8. What are the possible disadvantages and risks of taking part?

Any foreseeable discomforts, disadvantages and risks will be explained. The organizers/researchers will make known to the participants any predictable detriment arising from the proposed research activity. Furthermore, participants should be informed that any unexpected discomforts, disadvantages and risks to them, which arises during the research, should be brought immediately to the attention of the organizers.

9. What are the possible benefits for the participants of taking part?

The foreseen benefits to the participants that are expected will be presented. If there are no intended benefit to the participant from taking part in the project/activity this will be also clearly stated. The benefits for the project from the participant involvement will be explained as well.

Example opening sentence:

'Whilst there are no immediate benefits for those people participating in the project, it is hoped that this work will….'

10. What if something goes wrong?

The participants will be informed about how complaints will be handled and what redress may be available (i.e. what the process is). Handling complaints from participants regarding their treatment by researchers should be distinguished from events when something serious occurring during or following their participation in the project (e.g. a reportable serious adverse event). In the first instance the participants will know which member of the research project they should contact if they wish to raise a complaint. However, the participants will be also informed that should they feel their compliant has not been handled to their satisfaction, they can contact their DiverSea contact point or the Leader of WP 7.

11. Will the participation in the project/activity be kept confidential?

Participant's permission will be obtained to allow restricted access to information collected about them during the course of the project. They will be informed that all information will be kept strictly confidential and will know what measures will be taken to ensure this.

Example paragraph:

'All the information that we collect about you during the course of the research / interview/ project will be kept strictly confidential. You will not be able to be identified in any ensuing reports or publications.'

DiverSea researchers are aware that they are responsible for ensuring that when collecting or using data, they are not contravening the legal or regulatory requirements in Bulgaria, Sofia University and DiverSea.

12. What will happen to the results of the research project/activity?

It will be explained what will happen to the results of the research/activity (i.e. when the results are likely to be published, where they can obtain a copy of the published results, whether they be told which part of the project they were involved in) and that participants will not be identified in any report or publication.

Depending on the nature of the project, a statement might be included when needed, indicating that the data collected during the course of the project might be used for additional or subsequent research. In such case clear explanation of the purposes of the subsequent research will be provided.

13. Who is organising and funding the project activities?

The funding programme, organisation or company that is sponsoring or funding the research will be presented.

14. Contact for further information

A contact point for further information will be provided. This will include the name, address and telephone number, email.

15. Others

The information sheet will state that the participant will be given a copy of the information sheet, together with a signed consent form to keep and remember.

When the activity provisions producing recorded media additional questions will be included

16. Will the participants be recorded and how will the recorded media be used?

Participant's permission to record their activities on audio or video media will be obtained in advance. It will be ensured that there is a clear understanding as to how these recorded media will be used. For example, if a presentation is taking place, it will not be published or broadcasted, showed in public, or deposited in an archive without the presenter's permission. Storage and eventual disposal of interview recordings which contain sensitive material will be also covered.

Example paragraph:

'The audio and/or video recordings of your activities made during this interview/ project will be used only for analysis and for illustration in conference presentations and lectures. No other use will be made of them without your written permission, and no one outside the project will be allowed access to the original recordings.'

If the recordings are to be used in a publication or broadcasted or deposited in archive, a separate release form for each item used will be prepared and signed.