Funded by the European Union

SEAGLOW	
Project number:	101157477
Project name:	Sustainable Energy Applications for Green and Low-impact Operation of small-scale fishing boats in the Baltic and North Sea basins (SEAGLOW)
Topic:	HORIZON-MISS-2023-OCEAN-01-05
Type of action:	HORIZON
Starting date of action:	1 May 2024
Project duration:	48 months
Project end date:	30.04.2028
Deliverable number:	D7.2
Deliverable title:	Data Management Plan
Document version:	Ver1
WP number:	WP7
Lead beneficiary:	P1 - NDEU
Main author(s):	Steffen Helledie and Emilie Marie Nielsen – P1 NDEU
Internal reviewers:	
Nature of deliverable:	R
Dissemination level:	PU
Delivery date from Annex 1:	31-10-2024
Actual delivery date:	22-10-2024

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.

Version log

Version	Date	Released by	Nature of change
V1	22/10/2024	NDEU	1st draft

Executive Summary

This Data Management Plan (DMP) describes the data management life cycle of the data to be collected, generated by the project, and laying out the approach for their sound and FAIR management. The DMP will evolve during the lifespan of the project as a living document and provide details on the data as well as their management, including what type of data, how the data will be collected, shared, handled, preserved, what kind of standards will be applied, etc., ensuring transparency.

Abbreviations

D	Deliverable
DMP	Data Management Plan
EEAB	External Expert Advisory Board
EC	European Commission
PMH	Project Management Handbook
WP	Work Package
WT	Work Task

Contents

1 Introduction

The purpose of the SEAGLOW Data Management Plan (DMP) is to contribute to good data handling during and after the project's existence. It will describe what data the centre will generate, how the datasets will be stored and handled, whether and how data will be exploited or made accessible for verification and reuse, and how data will be curated and preserved.

2 Scope

This DMP applies to data that may be shared with the SEAGLOW consortium, the associated EEAB, the wider scientific community, or society at large. Datasets which are only accessible to one partner will be managed by that partner as they see fit and will not be covered by the DMP.

3 Data overview

In this section, we give a high-level overview over the data that will be collected, generated, and used in SEAGLOW. Due to the diversity of the project's research activities, there may arise cases of data that do not fit this description, which will be handled on a case-by-case basis and added to this DMP when relevant.

3.1 Purpose of data collection

The scope of SEAGLOW is to generate new knowledge, test new technologies, share knowledge between researchers, and inform the public. To achieve this, the project will collect and generate new data in addition to using existing datasets.

The purposes for collecting or generating data are:

- Research
- Method development, testing, and validation
- Technology development, testing and validation
- Knowledge sharing between researchers and fishing/maritime community

3.2 Types of data

In SEAGLOW, we expect to create the following types of data:

- Measurements from field experiments
- Measurements from laboratory experiments
- Environmental measurements

- Simulation results
- Research protocols
- Interviews
- Technical data
- Software source code

Data will be organised in datasets according to content and format.

3.3 Data formats

To the extent reasonable, we will use open and/or widely accepted data formats. Examples of such formats include, but are not limited to:

Documents (e.g., reports and publications): PDF/A, plain text, DOCX, ODF, LaTeX, HTML

Spreadsheets: XLSX, ODS

Tabular/array data: CSV, NetCDF

Images: PNG, JPEG, SVG

Audio: MP3, WAV, Ogg Vorbis, FLAC

Video: MP4, AVI, Matroska

Archives: ZIP, TAR (uncompressed, GZip, XZ, BZip2, Zstd)

Structured data (e.g., databases): SQL, JSON, XML, Parquet, HDF5

Source code: TXT (plain text for scripts), IPYNB (Jupyter notebooks)

By prioritizing these formats, we ensure long-term accessibility and interoperability of our data across platforms and research environments.

4 Data management

In SEAGLOW, data management will follow the principle: “as open as possible, as closed as necessary.” Our goal is to maximize the accessibility and reuse of the research data generated by the project. Whenever data can be shared openly and could potentially benefit others, we will make it publicly available. These datasets will be referred to as *open data*.

However, certain datasets cannot be made openly available due to reasons such as commercial sensitivity, the inclusion of personal information, or intellectual property rights (IPR) restrictions. These datasets will be classified as *closed data*. Access to closed data may be limited to project participants, specific companies, or even individuals with specific permissions.

This section outlines the management approach for both open and closed data. In Section 5, we detail additional plans to maximize the reuse of our open data.

4.1 Security classification

Each dataset will be classified according to the sensitivity of its contents and its need for protection. We will use four classes, in order of increasing security:[1]

Security class	**Description**
Unrestricted	The data can be freely distributed or used for publication purposes. Disclosure of the information does not harm anyone or anything. These datasets contain no sensitive or confidential information and can be openly distributed without restrictions.
Internal	The information may be accessed by all centre partners. Access to specific external parties may be granted in a controlled manner. It may also include preliminary findings or internal reports.
Confidential	The data may only be accessed by specific partners or individuals who have a legitimate need for the information to perform assigned tasks. Disclosure of the information to unauthorised parties could harm public interests, individuals, or centre partners. This classification could apply to datasets involving commercial information or data subject to contractual agreements.
Strictly confidential	The data may only be accessed by specific individuals who have a legitimate need for the

[1] These classes are a mix of the definitions used by NDEU, SINTEF and NTNU, with some adaptations for a project like SEAGLOW. The definitions of the classes “confidential” and “strictly confidential” in terms of the potential harm from unintended disclosure are borrowed from the Protection directive §4.

	information to perform assigned tasks. Disclosure of the information to unauthorised parties could cause significant harm to public interests, individuals, or centre partners. Data with the highest sensitivity, such as datasets containing personal information, commercially sensitive data, or subject to strict legal, regulatory, or IPR restrictions.

The last three may be referred to collectively as “restricted”.

Note that the security classes are not statements about usage rights and conditions. For example, publication of an “unrestricted” data set will still usually require consent from the data owner(s) and proper attribution.

Also note that the security classification is separate from the open/closed distinction made earlier, which refers to the actual dissemination level. In particular, all open data must be unrestricted, but not all unrestricted data need be open. One example is data which do not need to be protected, but which also have no interest for anyone else; another is data which will eventually be published openly, but not until they have undergone quality assurance.

4.2 Storage

Data will primarily be curated and stored by the partner responsible for generating them, following their internal data management procedures.

As the host institution, NDEU provides a Microsoft Teams channel, available to all partners, as a shared workspace and centralized storage area. This platform can be used to store and share data files with other project partners. To set up a new directory for storing data files, please contact the project’s Data Manager (see Section 7.2).

All files and data stored in NDEU's Microsoft Teams environment comply with EU data protection regulations, ensuring the security and privacy of all shared information.

4.3 Data sharing and access control

When one partner is to share data with another, they must do so using standard protocols and a level of encryption and access control which is appropriate for the data’s security classification. For restricted data, examples include:

- HTTPS with access control (e.g. a password protected web site)
- SSH

Use of email or messaging applications for data transfer should be avoided for confidential data and may not be used at all for strictly confidential data.

Access to the NDEU Microsoft Teams folders is controlled by TLS encryption and 2-factor authentication.

4.4 Archiving

All data underlying or comprising project results, as defined by the SEAGLOW Consortium Agreement, will be retained for a minimum of 10 years, unless legislation (e.g., the Personal Data Act or GDPR) prohibits it. Responsibility for the storage of these data lies with the partner(s) who own them.

Datasets stored in NDEU's Microsoft Teams folders will also be retained for at least 10 years after the conclusion of SEAGLOW, in accordance with NDEU's IT policy, unless restricted by applicable legislation.

Datasets containing personal data will be anonymized or deleted by the end of the project, ensuring compliance with relevant privacy laws.

Open data will be archived in public research data repositories, as outlined in the next section, and will be accessible for a minimum of 10 years.

5 FAIR data

This section only applies to open data.

To maximize the likelihood of reuse, we will ensure that all open data adhere to the FAIR principles: *Findable, Accessible, Interoperable, and Reusable*. To achieve this, the data will be published in repositories equipped to support FAIR compliance. Subsections 5.1–5.4 provide detailed explanations of how each principle is put into practice.

Our default repository for most cases will be the general-purpose **Zenodo** repository, which meets the requirements for FAIR data. However, we may also utilize other repositories in specific cases where they offer specialized support for certain data types or better facilitate the FAIR principles for the target audience. For example, the **PANGAEA** information system will be considered for georeferenced data from earth system research. We will prioritize repositories that are certified for best practices in data preservation and access.

In cases where the standard publication platform does not fully support FAIR principles, we will take supplementary actions, such as publishing metadata, or if necessary, entire copies of datasets on

additional platforms that ensure FAIR compliance. For example, when publishing open-source software, platforms like GitHub may be used, which are integrated with Zenodo to enable seamless archiving and citation for academic purposes.

5.1 Making data findable

Each open dataset will be assigned a unique *Digital Object Identifier* (DOI), ensuring persistent and easy identification. For updated versions of datasets, DOI versioning will be applied, providing unique identifiers for each version.

All datasets will be accompanied by comprehensive metadata that will include, at a minimum:

- DOI
- Creators and their affiliations
- Title
- Description or abstract
- Keywords
- Project name (SEAGLOW)
- Grant information (Horizon Europe project number: 101157477)
- License
- Language

The data repository will offer the necessary facilities for entering, storing, presenting, and searching this metadata, ensuring that the datasets are easily discoverable by others.

5.2 Making data accessible

Each open dataset will be downloadable using standard protocols (HTTP, FTP, etc.). Its DOI will lead directly to a downloadable resource, or to a resource that clearly shows how the data may be obtained. No authentication will be required.

Metadata will also be downloadable in a structured, machine-readable format. Usually, the repository will provide the facilities to export metadata to standard formats like Dublin Core.

The metadata will also be included within the dataset itself, to the extent that this is supported by the data format.

Data and metadata will remain accessible for a period of at least 10 years after publication.

5.3 Making data interoperable

Data will be published in open and/or widely accepted data formats. (See section 3.3 for examples.)

We will prefer the use of formats that have rich metadata facilities, in particular those that enable the creation of self-descriptive datasets (for example NetCDF).

5.4 Making data reusable

We will publish data under well-known and widely accepted licenses that permit wide reuse with minimal restrictions.

For most types of data, we will use Creative Commons licenses, with CC BY as the default choice.

For open-source software, we will use Open-Source Initiative approved software licenses (for example the MIT License).

6 Ethical and legal aspects

6.1 Research ethics

The work in SEAGLOW, including data management, will be conducted in accordance with the general research ethics guidelines:

- General guidelines for research ethics
- Guidelines for research ethics in science and technology
- Guidelines for research ethics in the social sciences and the humanities

6.2 IPR

Intellectual property rights are governed by the SEAGLOW Consortium Agreement.

6.3 Confidentiality

Confidentiality is governed by the SEAGLOW Consortium Agreement.

7 Allocation of resources

7.1 Costs

The costs of data storage, including long-term archiving, will be borne by the partner responsible for storing the data, and not covered by the project budget.

For long-term preservation of open data, we will use free-of-charge data repositories like Zenodo.

The costs of data management activities are limited to project management costs and will be covered by allocated resources in the project budget.

Other resources needed to support reuse of data after the project ends will be solved on a case-by-case basis.

7.2 Data manager

The ultimate responsibility for data management in SEAGLOW lies with the Steering Group. On a day-to-day basis, this responsibility is delegated to the *Data manager*, which is currently Emilie Marie Nielsen (P1 - NDEU).