TWInning Laboratory for an Innovative Global Hub To Explore the Deep

Deliverable 1.2

Data management plan

Diane Esson[1], Deise Faria[1], Jan Dierking[2], João Canning-Clode[1], Ana Širović[3]
(1) ARDITI, (2) GEOMAR, (3) NTNU

Deliverable number and name	1.2 Data management plan
Work package	1: Planning and definitions
Related task in DoA	1.2 Data management planning
Due date	31 March 2025
Submission date	27 March 2025
Responsible organization	ARDITI
Author name(s)	Diane Esson, Deise Faria, Jan Dierking, João Canning-Clode, Ana Širović
Dissemination level	PU (public)

Revision history

Version	Date	Modified by	Comments
1	5-Feb-2025	DE	First draft
2	10-Feb-2025	DF	Edits to draft of the deliverable
3	25-Feb-2025	DE	Full draft, reviewed by Data Managers, ready for review by consortium leadership
4	12-Mar-2025	JCC, DE	Revisions
5	12-Mar-2025	JD	Revisions
6	19-Mar-2025	AŠ	Revisions
7	26-Mar-2025	DE	Final version

Disclaimer

Views and opinions expressed in this deliverable are those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible.

Table of contents

1.   Overview

The Data Management Plan (DMP) summarizes the data infrastructure and management procedures to be used within the TWILIGHTED project. As TWILIGHTED is a capacity-building project, this DMP also reviews training to help ARDITI develop its data management systems and strategies in future.

2.   Data Summary

2.1.   Project introduction

TWILIGHTED is a Horizon Europe Twinning project that seeks to revolutionize deep-sea research, development and innovation (RD&I) in Portugal, ultimately benefiting all of Europe. Collaborating closely with GEOMAR (Germany) and NTNU (Norway), TWILIGHTED aims to transform ARDITI (Madeira, Portugal) into a global hub for deep-sea RD&I. Focusing on the Mesophotic Zone (40-200m) and the Twilight Zone (200-1,000m), and capitalizing on Madeira's unique proximity to deep waters, TWILIGHTED will help accelerate our global understanding of the ocean and its essential role in sustaining life on earth.

TWILIGHTED's key objectives are to: (1) Collaborate across research institutes in Europe, (2) Elevate the research profile of Portugal and especially the European Outermost Region of Madeira, (3) Innovate low-cost alternatives to state-of-the-art deep-sea research technologies, (4) Democratize deep-sea research, (5) Globalize deep-sea RD&I and (6) Share ocean science across stakeholders.

To achieve its objectives, TWILIGHTED will adopt state-of-the-art approaches to training and networking. Development activities include staff exchanges, expert visits, training schools, joint research missions, the International Twilighted Conference and novel cross-sector workshops stimulating creativity in solving the ocean's greatest challenges (the Impossible Things Workshops). Such capacity-building will not only catalyze deep-sea RD&I in Madeira but also ensure a lasting impact on the deep-sea scientific landscape, sustainable ocean policy, and Madeira's socioeconomic development.

2.2.   Project-relevant data

TWILIGHTED will produce considerable capacity-building knowledge from different sources, with diverse data formats, sensitivity and use-cases. All data will be stored in line with the requirements of the General Data Protection Regulation (GDPR) and data will be curated by partners under the supervision of the Coordinator and TWILIGHTED Data Managers. Data will be held on secure servers at ARDITI's internal cloud and in the Google Drive cloud.

Data expected to be generated and used during this project includes (1) new scientific data collected during training expeditions, training workshops or technology trials, in the form of spreadsheets, images, videos, acoustic recordings, and specimen vouchers (2) shared scientific

data from project partners, primarily background data for training purposes and (3) personal data collected for the purposes of coordination, recruitment, event attendance or project dissemination or communication.

2.3. Purpose of the Data Management Plan

The Data Management Plan (DMP) serves as a framework for analyzing the Consortium's data management policy on project data. Covering the entire data lifecycle, the DMP outlines the types of data generated or collected, the standards employed, preservation methods, and the extent to which datasets are shared for verification or reuse. Additionally, it reflects the Consortium's agreements on data management, ensuring alignment with exploitation and intellectual property rights (IPR) requirements.

As a dynamic document, the DMP evolves throughout the project, particularly in response to significant changes such as dataset updates or modifications in Consortium policies. This document represents the initial version of the DMP, developed by month six (M6) of the project, in accordance with the Grant Agreement (GA). It provides an overview of the key datasets expected to be produced and specifies the conditions associated with their use. This version has been prepared in compliance with the European Commission's guidelines for Horizon Europe (HEU) projects. As a working document, an updated version will be provided at M15, and a final version produced by the end of the project. ARDITI will be responsible for overseeing that all partners manage data in accordance with the DMP and meet appropriate local legal and ethical standards.

While adhering to open-access requirements, the TWILIGHTED project also prioritizes the privacy of individual participants. To safeguard personal data, only fully anonymized information may be shared publicly, ensuring that individuals cannot be identified. Non-anonymized data is stored internally and is accessible only to the designated person or team responsible for the specific activity or research. In cases where non-anonymized data is intended for public use, such as the names of R&I experts or recordings/photos from project events, explicit consent should be obtained from the relevant individuals before publication.

3. Data formats generated or collected in TWILIGHTED

Scientific data may be generated during this capacity-building project during training expeditions or workshops or technology trials, and could take the form of numerical, acoustic, visual, genetic, taxa or descriptive data. The purpose of this data is to aid practical training in deep-sea research techniques and the development of low-cost research technologies.

Personal data collected during this project may include names and email addresses for project mailing lists, scientific manuscripts, presentations and activity attendance (e.g. for meetings/workshops, seminars, and conference sessions). Information on diversity metrics will also be collected for reporting purposes, at the discretion of the individual. More extensive personal data will also be collected for recruitment purposes (i.e. to identify candidates to

support in external PhD or postdoctoral fellowship applications). This data will take the form of resumés and cover letters reviewing educational and career achievements. The purpose of the personal data collected in TWILIGHTED is to provide project partners with the necessary information to carry out planned tasks, to report project activities to the European Commission (EC), to recruit talent and to disseminate information. As such, this personal data directly contributes to the objectives of the project. Data will be useful mainly to the EC, stakeholders, and project participants/partners, with the exception of communication and dissemination of information aimed at a more general audience. All personnel data will be stored in line with GDPR (see section 6).

4.   Re-use of existing data

Data to be reused is declared under the "Background Included" section in the CA. Some restrictions in the use of data may apply related to the declared background in the CA (Attachment 1: Background included).

5.   Origin of data

Scientific data is expected to arise from TWILIGHTED training schools and cruises, staff exchanges, expert visits and innovation workshops. Internal personal data will be collected for attendance at TWILIGHTED activities and involvement in TWILIGHTED scientific publications. External personal data will be collected for attendance at the Impossible Things Workshop and International Twilighted Conference and for any publications (e.g. white papers) related to these activities; external personal data will also be collected for any TWILIGHTED-related personnel recruitment, including applications to PhD or post-doctoral fellowship calls.

6.   Treatment of personal data

In accordance with the GA, project members must process personal data in compliance with the applicable EU, international and national law on data protection, and in particular, Regulation 2016/679. As such, personal data must be:

- processed lawfully, fairly and in a transparent manner in relation to data subjects
- collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes
- adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed
- accurate and, where necessary, kept up to date
- kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data is processed and
- processed in a manner that ensures appropriate security of the data.

Personal data collected for TWILIGHTED activities, publications or fellowship applications will be stored in accordance with GDPR on ARDITI's internal cloud and/or the Google Drive cloud

for a maximum period of 90 days after the completion of the activity or submission of the publication or application, after which this data will be digitally deleted from all cloud platforms and any institutional servers and any printed copies will be shredded. Only project members directly involved in the project activity that involves each personal dataset will have access to this data.

7.   FAIR treatment of scientific data

In accordance with FAIR principles, all scientific data generated by the TWILIGHTED project will be made publicly accessible and findable. In line with best practices, TWILIGHTED will collect smaller datasets in controlled settings and continuous longitudinal datasets. All data will be subject to cleaning to make the data FAIR compatible.

All datasets produced by TWILIGHTED activities and uploaded to database platforms are to be recorded in the 'Data Management Database' Google Sheet (Figure 7a), found in the TWILIGHTED Google Drive > WPs > WP 1/2 > Reporting. Each entry should be verified by a TWILIGHTED Data Manager.
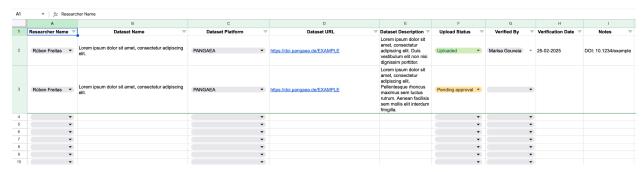

Figure 7a. Data management database screenshot

7.1.   Making scientific data accessible

All project data, metadata, or appropriate outputs will be made publicly available and as soon as possible following generation of that data. Immediate accessibility may not be possible in some cases where moratoria exist (i.e. before peer-reviewed publications have been published). In cases when open accessibility to raw data is not possible (e.g., due to national security concerns), metadata and information regarding possible ways of accessing raw data will be made public. Means of data and metadata publication include the project website, partner websites, and long-term public repositories such as the EU Marine Observation and Data Network (EMODNet), the EU Open Science Cloud portal (ESOC), and other relevant platforms (e.g., OBIS, GBIF, PANGAEA, Zenodo). For example, all final numerical datasets will be stored and made publicly available on PANGAEA. If collected, genetic data will be archived in GenBank.

Third parties will be granted access, mining, exploitation, reproduction and dissemination of scientific data free of charge, beyond any costs of sending physical data, with due consideration given to any potential intellectual property (IP) exceptions in accordance with the Consortium Agreement (CA), which will serve as the basis for project IP protection guidelines, or based on any national security concerns.

Repository

As listed above, public repositories including the EU Marine Observation and Data Network (EMODNet), the EU Open Science Cloud portal (ESOC), and other relevant platforms (e.g., OBIS, GBIF, PANGAEA, Zenodo) will be used to make data and metadata accessible. The project will also upload on the ARDITI open access repository any scientific data from the project that can be shared openly. In all cases, data will abide by internationally accepted and accessible file formats, in line with FAIR principles.

7.2. Making scientific data findable, including provisions for metadata

Open-access publications

Where relevant, scientific data will be compiled and submitted for publication in open-access articles. These articles will be included in scientific paper search engines ResearchGate and Google Scholar and also published on the TWILIGHTED website and social media platforms to aid findability.

Metadata

Creating metadata related to the data to be created is not always applicable. The creation of metadata will be analysed on a case-by-case basis when making data available through public repositories. Where applicable, metadata of deposited data will be open under a Creative Common Public Domain Dedication (CC 0) or equivalent (to the extent legitimate interests or constraints are safeguarded), in line with the FAIR principles, as mentioned in the Grant Agreement. Metadata will be mainly in PDF, DOC and XLSX formats by default. Where applicable, the metadata will include persistent identifiers for related publications and other research outputs.

7.3. Scientific data re-use

New scientific datasets are not intended for reuse. However, if applicable, standard reuse licenses will be implemented in accordance with the obligations outlined in the GA (e.g., Creative Commons Public Domain Dedication (CC0) or an equivalent license).

8. Other research outputs

Online access to research outputs is provided free of charge to the end-user, such as scientific publications, data, metadata or other engineered results and processes such as software,

algorithms, protocols, models, workflows and electronic notebooks, complying with open science standards and conditions defined in Article 17 of the GA.

9.   Allocation of Resources

Each TWILIGHTED partner has a budget for Open Access publications, ranging from €2,500 to €15,000 to augment existing institutional support for Open Access publications at ARDITI, GEOMAR and NTNU. ARDITI has allocated budget for five publications, GEOMAR has allocated budget for two publications and NTNU has allocated budget for one publication. All partners also have Online Knowledge Library (B-on or equivalent) subscriptions, which guarantee open access status for publications in its journals.

10.   Data manager responsibilities and training

As part of this capacity-building project, TWILIGHTED will support data management training and career development at ARDITI. For this, ARDITI has named two data managers, Rúben Freitas and Marisa Gouveia, to attend training courses and support ARDITI in developing best-practice in data management. To date, such training includes the PANGEA Community Workshop on FAIR data publications within PANGEA (7-8 November 2024), an opportunity shared by GEOMAR. At the 2025 training school at GEOMAR (M3.1), there will also be a data management and communications training day (M8.1), which will include sample and data management and image data curation and management (see D3.1 for more details). Future opportunities will be designed or disseminated by TWILIGHTED partners, including online meetings with data managers at GEOMAR.

In addition to sharing data management training resources with TWILIGHTED members and ARDITI staff, TWILIGHTED's data managers will be responsible for supporting ARDITI researchers in designing appropriate databases and/or data-sharing platforms, offering guidance on cleaning data to make it FAIR compatible and confirming that data is uploaded to relevant repositories for any TWILIGHTED scientific data (see section 7).

Adherence to the DMP and fair and ethical treatment of data is the responsibility of all consortium leadership, with ultimate oversight from the project coordinator.

11.   Data security

Sensitive data will be available only internally through Google Drive with restricted access or via disk sharing. Open data will be made available via the project website, partner websites and long-term public repositories such as the EU Marine Observation and Data Network (EMODNet), the EU Open Science Cloud portal (ESOC), and other gateways (e.g. OBIS, GBIF, PANGAEA).

12. Ethics and informed consent

No ethics or legal issues apply to datasets to be created. Recording of personal data (like names and email addresses) will follow an opt-in procedure that articulates the data privacy statements of ARDITI, GEOMAR or NTNU as relevant, and this data will not be shared publicly. More information on the treatment of personal data can be found in the Ethics Plan, D1.3.