ColossuS

Multi-purpose unmanned vehicle swarm patrol intelligence platform for port security and resilience

D8.2 Data Management Plan – 1st Version

Q-PLAN

27/02/2026



Co-funded by the European Union

Document Information

Issued by:	Q-PLAN
Issue date:	27/02/2026
Due date:	28/02/2026
Work package leader:	Q-PLAN
Dissemination level:	PU (Public)

Document History

Version	Date	Modifications made by
0.1	11/02/2026	Q-PLAN – Initial draft
0.1	25/02/2026	ALL Partners – Input and Review
1.0	26/02/2026	Q-PLAN – Revisions based on partners input
1.0	27/02/2026	Q-PLAN – Final version available for submission

Authors

Names	Beneficiary
Tsolakis A., Boikou A.	Q-PLAN

In case you want any additional information, or you want to consult with the authors of this document, please send your inquiries to: boikou@qplan-intl.gr

Quality Reviewers

Name	Beneficiary
Zulmira Cunha	TV
Ricardo Peres	UNINOVA
Argiro Amidi, Panagiotis Karkazis	iNGS
Pavel Prautsch	ABB
Vitor Nogueiro	AASA

Disclaimer

Co-Funded by the European Union under GA no. 101225980. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or REA. Neither the European Union nor the granting authority can be held responsible for them.

© COLOSSUS Consortium, 2026

Reproduction is authorised provided the source is acknowledged.

Executive Summary

This document constitutes the initial version of the Data Management Plan (DMP) and has been elaborated as a deliverable (D8.2) in the framework of the COLOSSUS project. COLOSSUS presents an integrated solution for digital security, resilience, and maintenance of port infrastructures, providing real-time data access from autonomous vehicles and sensors. Developed in response to the increasing demand for maritime security, COLOSSUS aims to revolutionize the protection of EU maritime infrastructures, aligning with the updated EU Maritime Security Strategy (EUMSS 2023).

In this context, the initial version of the project's DMP sets out the overall methodological principles pertaining to the management of the data that will be collected, generated and/or re-used in the framework of COLOSSUS, safeguarding sound and ethical data management along the entire duration of the project. Moreover, it provides a first, yet still meaningful overview of COLOSSUS' data, as identified in this early stage of the project, along with information on the methodology pertaining to their management as well as to making them Findable, Accessible, Interoperable and Re-usable (FAIR).

The initial version of the DMP is the first of the three versions of COLOSSUS' DMP deliverables to be produced in the course of the project (D8.2 DMP – Version 1 delivered in M4 will be updated to D8.3 DMP - Version 2 in M12 and ultimately fixed as D9.1 DMP - Version 3 in M24). Along these lines, the DMP will be updated and further elaborated during the project to reflect an accurate, up-to-date and ultimately comprehensive plan for managing the data that will be collected, generated and/or re-used by the project across their entire life cycle, both during and after the completion of COLOSSUS.

Table of Contents

1. INTRODUCTION 8
1.1. SCOPE OF THE DELIVERABLE 8
1.2. STRUCTURE OF THE DELIVERABLE 9
2. DATA SUMMARY 10
2.1. PURPOSE OF DATA COLLECTION/GENERATION OR RE-USE AND ITS RELATION TO THE OBJECTIVES OF THE PROJECT 10
2.2. TYPES AND FORMATS OF COLLECTED/GENERATED OR RE-USED DATA 11
2.3. ORIGIN OF DATA AND RE-USE OF PRE-EXISTING DATA 16
2.4. EXPECTED SIZE OF DATA 17
2.5. DATA UTILITY 19
3. FAIR DATA 20
3.1. MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA 20
3.2. MAKING DATA ACCESSIBLE 24
3.3. MAKING DATA INTEROPERABLE 31
3.4. INCREASE DATA RE-USE 32
4. OTHER RESEARCH OUTPUTS 37
5. ALLOCATION OF RESOURCES 39
5.1. ESTIMATED COSTS FOR MAKING DATA FAIR 39
5.2. DATA MANAGEMENT RESPONSIBILITIES 40
6. DATA SECURITY 44
7. ETHICS AND OTHER ISSUES 46
8. CONCLUSIONS 48
ANNEXES 49
ANNEX I – PRIVACY POLICY 49
ANNEX II – INFORMED CONSENT FORM 54
ANNEX III – DATA SUBJECT REQUEST FORM 57
ANNEX III – RECORD OF PROCESSING ACTIVITIES 60

List of Figures

Figure 1: Typical DOI created by Zenodo. 24

List of Tables

Table 1. Terms and Definitions 7

Table 2. COLOSSUS' partners 8

Table 3 - Expected Size of Data 17

Table 4 - Data Utility 19

Table 5: Good practices for data anonymisation 25

Table 6: Data availability 26

Table 7: Data accessibility 28

Table 8: Dublin Core Metadata standard vocabulary 31

Table 9: Expected time that data will be made open through Zenodo 34

Table 10: Other research outputs 37

Table 11: Estimated costs for making data FAIR 39

Table 12: Data management responsibilities per data collected/generated under each WP 42

Table 13. Record of Processing Activities 60

List of Terms and Definitions

Table 1. Terms and Definitions

Abbr.	Definition	Abbr.	Definition
3D	Three-dimensional	IP	Internet Protocol
AASA	Arsenal do Alfeite	ISP	Internet Service Provider
AI	Artificial Intelligence	JSON	JavaScript Object Notation
AIoD	AI on Demand	KPI	Key Performance Indicators
API	Application Programming Interface	NAS	Network-Attached Storage
BSD	Berkeley Software Distribution	OAI	Open Archives Initiative
CERN	European Organization for Nuclear Research	OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
CESSDA	Consortium of European Social Science Data Archives	PC	Project Coordinator
DCMI	Dublin Core Metadata Initiative	PID	Persistent Identifier
DMP	Data Management Plan	QA	Quality Assurance
DoA	Description of Action	QC	Quality Control
DOI	Digital Object Identifiers	SAB	Security Advisory Board
DPA	Data Protection Authority	SDK	Software Development Kit
EC	European Commission	SSH	Secure Shell
EEA	European Economic Area	UAV	Unmanned Aerial Vehicle
EU	European Union	UGV	Unmanned Ground Vehicle
EUMSS	EU Maritime Security Strategy	UOPV	Unmanned Oceanic Patrol Vessel
FAIR	Findable, Accessible, Interoperable and Re-usable	URL	Uniform Resource Locator
GA	Grant Agreement	USV	Unmanned Surface Vessel
GDPR	General Data Protection Regulation	UV	Unmanned Vehicles
HTML	HyperText Markup Language	WP	Work Package
HTTPS	HyperText Transfer Protocol Secure	WPL	Work Package Leader
ISMS	Information Security Management System	WTL	Work Task Leader

1. Introduction

1.1. Scope of the deliverable

The current document represents the initial version of COLOSSUS' Data Management Plan (DMP), which has received funding from the European Union's Framework Programme for Research and Innovation Horizon Europe under Grant Agreement No 101225980.

COLOSSUS presents an integrated solution for digital security, resilience, and maintenance of port infrastructures, providing real-time data access from autonomous vehicles and sensors. Developed in response to the increasing demand for maritime security, COLOSSUS aims to revolutionize the protection of EU maritime infrastructures, aligning with the updated EU Maritime Security Strategy (EUMSS 2023).

The advanced COLOSSUS solutions will be demonstrated in an operational environment (integrated pilot system level) for iterative performance evaluation and validation. Starting with stress testing in a laboratory setting, the project will move towards real-world operational conditions in ports. Defined use cases will enable operational evaluation and validation across diverse scenarios at the Arsenal do Alfeite (AASA) naval base in Portugal.

To this end, the **consortium** of COLOSSUS brings together a complementary and interdisciplinary group of **6 partners across 3 different countries** within the EU, as presented in the **Table 2**.

Table 2. COLOSSUS' partners

Partner Role*	Partner No	Partner Name	Partner Short Name	Country
COO	1	TECNOVERITAS SERVICOS DE ENGENHARIAE SISTEMAS TECNOLOGICOS LDA	TV	Portugal
BEN	2	UNINOVA-INSTITUTO DE DESENVOLVIMENTO DE NOVAS TECNOLOGIAS-ASSOCIACAO	UNINOVA	Portugal
BEN	3	INTELLIGENT NEXT GENERATION SYSTEMS LYSEIS PLIROFORIKIS KAI EPIKOINONION SA	iNGS	Greece
BEN	4	Q-PLAN INTERNATIONAL ADVISORS PC	Q-PLAN	Greece
BEN	5	ABB SRO	ABB	Czechia
BEN	6	ARSENAL DO ALFEITE SA	AASA	Portugal

*COO: Coordinator, BE: Beneficiary

All partners of COLOSSUS' consortium adhere to sound data management principles in order to ensure that the meaningful data collected, generated and/or re-used throughout the duration of the project are well-managed, archived and preserved, in line with the structure and guidelines of the **Horizon Europe Data Management Plan Template**[1].

The methodology of COLOSSUS for data management builds on know-how, tools and templates that were developed internally by Q-PLAN as well as on good practices and templates from the literature (such as the Horizon Europe Data Management Plan Template). As in previous EU-funded projects, tailored modifications to the methodology were implemented for COLOSSUS as well, in order to comply with the GA conditions and the particularities of the project. Along these lines, this deliverable presents the adjusted methodology as it was

[1] https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/temp-form/report/data-management-plan_he_en.docx

further developed and applied in the context of COLOSSUS, as well as presents the results from its application during the project.

Along these lines, this initial version of the DMP aims to achieve the following objectives:

- Describe the data management lifecycle for the data to be collected, generated and/or re-used in the framework of COLOSSUS, serving as the key element of good data management.
- Outline the methodology employed to safeguard the sound management of the data collected, and/or generated as well as to make them Findable, Accessible, Interoperable and Re-usable (FAIR).
- Provide information on the data that will be collected, generated and/or re-used and the way in which it will be handled during and after the end of the project along with the standards applied to this end.
- Describe details on how the data will be made openly accessible and searchable to interested stakeholders as well as its curation and preservation.
- Address the management of any research outputs other than data in line with FAIR principles.
- Present information on the resources to be allocated so as to make data FAIR clearly identifying responsibilities pertaining to data management, while addressing data security and ethical aspects.

1.2. Structure of the deliverable

Taking the above into consideration, the report is structured as follows:

- **Chapter 1** provides introductory information about the DMP, the context in which it has been elaborated as well as its objectives and structure;
- **Chapter 2** presents a summary of the data to be collected/generated or re-used during the activities of COLOSSUS including its purpose as well as its types and formats. Additionally, it outlines its origin, expected volume and the stakeholders that may find it useful;
- **Chapter 3** describes the methodology that is applied in COLOSSUS in order to safeguard the effective management of data across their entire lifecycle, making it FAIR;
- **Chapter 4** presents the management of other research outputs that may be generated or re-used throughout COLOSSUS and provides sufficient details on making them FAIR;
- **Chapter 5** estimates the resources required for making the project's data FAIR, while also identifying data management responsibilities;
- **Chapter 6** outlines the data security strategy applied within the context of COLOSSUS along with the respective secure storage solutions employed;
- **Chapter 7** addresses ethical aspects as well as other relevant issues pertaining to the data collected/generated or re-used during the implementation of the project;
- **Chapter 8** concludes on the next steps foreseen in the framework of the project with respect to its data management plan.

Annexed in the document are (i) the project's Privacy Policy (Annex I), the templates for the (ii) Informed Consent Form (Annex II) and (iii) the Data Subject Request Form (Annex III) as well as (iv) the Record of Processing Activities (Annex IV), which will be used during the implementation of the project's activities to ensure compliance with relevant applicable EU and national regulation(s).

Note that the DMP is not a fixed document. It evolves during the lifespan of the project and will be further elaborated and updated at least twice more throughout the duration of COLOSSUS (i.e., as D8.3 at M12 and D9.1 at M24). Additional ad hoc updates may also be realised (if necessary), in order to include new data, better detail and/or reflect changes in the methodology or other aspects relevant to their management (such as costs for making data FAIR, size of data, etc.), changes in consortium policies and plans or other potential external factors. Q-PLAN is responsible for the elaboration of the DMP and with the support of all partners will update and enrich it when required.

2. Data Summary

COLOSSUS will generally collect/generate or re-use meaningful non-sensitive data that do not fall into any special categories of personal data as those are described within the General Data Protection Regulation (GDPR). Sensitive, though non-personal, data are expected to be collected/generated during the use case demonstration but will not be circulated to the consortium prior to their anonymisation or pseudonymisation. These data may be quantitative, qualitative or a blend of those in nature and will be analysed from a range of methodological perspectives with a view to producing insights that will successfully feed COLOSSUS' activities, enable us to deliver evidence-based results and ultimately achieve the objectives of the project. With that in mind, the second chapter of the DMP starts by explaining the purpose for which this data will be collected/generated and how it relates with COLOSSUS. It proceeds by describing the different types and formats of these data as well as their origin and expected volume, before concluding with an overview of potential stakeholders for whom they may prove useful for re-use.

2.1. Purpose of data collection/generation or re-use and its relation to the objectives of the project

In order to successfully meet its objectives and ensure the production of evidence-based results, COLOSSUS entails several activities during which data will be collected/generated or re-used. The purpose for which these data are collected/generated or re-used is interrelated with the objective of the activity during which they are produced.

In particular, these activities along with their objectives in the framework of COLOSSUS are as follows:

- **Market analysis for the identification of value chains and customer needs for the COLOSSUS solution**, in order to extract market-driven requirements for the design of the COLOSSUS system.
- **Identification of regulatory, ethics and security standards requirements** for driving compliance by-design. This includes the **Z-Inspection® Co-design / Co-assessment process,** in order to initially provide requirements towards a more trustworthy AI implementation, and later on assessing the trustworthiness of the overall COLOSSUS AI system.
- **Co-design of the COLOSSUS requirements and system architecture**, in order to deliver a product market-fit architecture that addresses not only technical requirements but also market, regulatory, ethics, and security standards requirements.
- **Pilot Use Case Scenarios Planning**, in order to elaborate the preparatory and deployment steps for a series of real-world scenarios under operational conditions (TRL7), that will provide the necessary context to evaluate and validate COLOSSUS' mature solutions.
- **Replication guidelines,** including documentation of best practices for implementation across diverse geographical and operational contexts, in order to support small innovators access the civil security market.
- **Policy recommendations,** analysing regulatory, technical, and operational barriers faced by SMEs wanting to access the civil security market, thus enhancing cooperation between public buyers and small supply market actors.
- **Business development** which includes business modelling, business planning and engagement with industrial actors in order to ensure the effective and viable exploitation of the COLOSSUS results, while securing customer leads.
- **Collaboration and Synergies with relevant projects and Initiatives**, in order to coordinate, develop and benefit from synergies with other relevant EU initiatives.
- **Setup and Operation of the Security Advisory Board (SAB).** The SAB comprises of experts in diverse domains who provide strategic guidance in key stages of the project, contributing with their expertise and representing the views and interests of their stakeholder communities in order to better align the results of the project with them.

- **Unmanned Surface Vessel (USV) data collection, processing and decision making**. xxx
- **Unmanned Aerial Vehicles (UAVs) and Unmanned Ground Vehicles (UGVs) data collection, processing and decision making.** xxx
- **Secure interfaces for port security information exchange**, in order to manage data from other COLOSSUS components, facilitating seamless, interoperable data exchange across maritime security stakeholders, including port authorities and coast guards.
- **Multi-modal data-efficient harmonisation**.
- **Intelligent Swarm navigation for diverse Unmanned Vehicles (UV) types.**
- **3D (three-dimensional) reconstruction and modelling for underwater inspection**.
- **Integration lab testing and fine tuning**, to ensure the correct acquisition, storage and analysis of the preliminary data according to the requirements and compliance related to security data and EU requirements.
- **Lab stress-testing and technology validation,** in order to adequately test the integrated solution and ensure its performance by passing industrial-grade testing procedures.
- **COLOSSUS Web Platform,** providing a uniform command and control platform for the handling the COLOSSUS assets and generated data/information in real-time operations.
- **Pilot Preparation and Deployment** in AASA under diverse operational conditions and use cases, to test its functionality, performance and resilience.
- **Technical evaluation and market validation,** involving a technical evaluation of COLOSSUS' use case performance and a market-fit validation of COLOSSUS's products based on end-users' feedback.
- **Monitoring and assessment of the dissemination, communication and stakeholder engagement activities** of COLOSSUS with a view to measuring their results and impact, fine-tune COLOSSUS' Dissemination and Communication Plan, as well as fulfil the project's reporting requirements towards the Commission.

The following section provides further details on the different types and formats of data collected/generated or re-used during the project's activities.

2.2. Types and formats of collected/generated or re-used data

2.2.1. Data collected/generated through direct input methods

Direct input methods, under the scope of COLOSSUS, involve methodologies for collecting data through desk research and interactions between consortium partners and external stakeholders, with the latter providing data to the former. Along these lines, external stakeholders undertake the role of a data subject that is a natural person whose personal data are being processed[2]. In particular, the identification and selection of suitable data subjects are based on purposeful sampling according to which, external stakeholders are identified and selected by consortium partners based on their relevance to COLOSSUS' subject area and the objectives of the respective activity for which data is collected. In this context, quantitative and qualitative data will be collected/generated during COLOSSUS[3]:

- **Quantitative data** are numerical and acquired through counting or measuring. Examples of quantitative data are the yearly turnovers of a business, the hourly compensation of a worker, the number of SMEs in Europe, etc. These data may be represented by ordinal, interval or ratio scales and lend themselves to statistical manipulation.

[2] 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, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679.
[3] Neuman, W. L. (2014). Social research methods: Qualitative and quantitative approaches. Boston: Pearson.

- **Qualitative data,** sometimes referred to as categorical data, is data that can be arranged into categories based on physical traits, gender, colours or anything that does not have a number associated with it. Moreover, written documents, interviews, and various forms of in-field observation are all sources of qualitative data. Examples of qualitative data are the preferences of learning, skillsets, country of origin, etc.

Additional details with respect to the different types and formats of data that will be collected through direct input methods under the frame of COLOSSUS are provided below.

1) Market analysis

The data to be collected/generated in the framework of this activity fall under Tasks 1.1 and 7.3 (Q-PLAN) and will be qualitative in nature. They will be used to analyse the current market, relevant to COLOSSUS. Data will be collected through desk research, semi structured interviews with key experts and online surveys. The data storage formats will include text documents, spreadsheets, images, presentations (.docx, .xlsx, .csv, .pptx, .jpeg, .png, .pdf).

2) Identification of regulatory, ethics and security standards requirements

The data to be collected/generated in the framework of this activity fall under Task 1.2 (iNGS) and will be qualitative in nature. They will be used to define the regulatory, ethical and security standards context within which the COLOSSUS system needs to be implemented and comply with. Data will be collected through desk research, and interaction with key experts following the Z-Inspection® process for Trustworthy AI. The data storage formats will include text documents and presentations (.docx, .pptx, .jpeg, .png, .pdf).

3) Co-design of the COLOSSUS requirements and system architecture

The data to be collected/generated in the framework of this activity fall under Task 1.3 (TV) and will be qualitative in nature. They will be used to identify and define the functional and non-functional requirements for the COLOSSUS solution as well as the detailed system architecture of the COLOSSUS system, covering in detail all components/users and their in-between interactions. Data will be collected through co-design workshops among partners and the SAB. The data storage formats will include text documents, spreadsheets, images, presentations (.docx, .xlsx, .csv, .pptx, .jpeg, .png, .pdf).

4) Pilot Use Case Scenarios Planning

The data to be collected/generated in the framework of this activity fall under Task 1.4 (AASA) and will be qualitative and quantitative in nature. They will be used to plan the pilot use cases as well as to better capture the current status of their operational context, defining and quantifying clear metrics to measure the performance of the COLOSSUS System. Data will be collected through co-design workshops among partners and the SAB. The data storage formats will include text documents, spreadsheets, images, presentations (.docx, .xlsx, .csv, .pptx, .jpeg, .png, .pdf).

5) Replication guidelines

The data to be collected/generated in the framework of this activity fall under Task 5.3 (UNINOVA) and will be qualitative in nature. They will be used to offer practical recommendations for small innovators/SMEs on how to access the civil security market. Data will be collected through desk research and workshops with partners to capture and document best practices for implementation in diverse geographical and operational contexts, addressing critical port security and resilience challenges. The data storage formats will include text documents, spreadsheets, images, presentations (.docx, .xlsx, .csv, .pptx, .jpeg, .png, .pdf).

6) Policy recommendations

The data to be collected/generated in the framework of this activity fall under Task 5.3 (UNINOVA) and will be qualitative in nature. They will be used to recommend policy measures to address regulatory, technical, and operational barriers faced by SMEs wanting to access the civil security market, and how to enhance cooperation between public buyers and small supply market actors. Data will be collected through desk research and workshops with partners, SAB members and key external experts. The data storage formats will include text documents and presentations (.docx, .pptx, .pdf).

7) Business development

The data to be collected/generated in the framework of this activity fall under Tasks 6.3 and 7.3 (Q-PLAN) and will be qualitative in nature. They will be used to plan and implement the business/go-to-market strategy for the COLOSSUS products and to secure customer leads. Data will be collected through desk research, workshops and interaction with industrial actors (within and outside the consortium and the SAB). The data storage formats will include text documents, spreadsheets, images, presentations (.docx, .xlsx, .csv, .pptx, .jpeg, .png, .pdf).

8) Collaboration and Synergies with relevant projects and Initiatives

The data to be collected/generated in the framework of this activity fall under Tasks 6.4 and 7.4 (Q-PLAN) and will be qualitative in nature. The data to be collected include, but are not limited to photos, number of event participants and presentations, meeting minutes, past knowledge and good practices deriving from the partners' synergetic actions with relevant EU initiatives and collaborations with sister projects. The data storage formats will include text documents, spreadsheets, images and presentations (.docx, .xlsx, .csv, .ppt, .jpeg, .png, .pdf). Additionally, some information will be available online on COLOSSUS' and collaboration partners' websites and social media.

9) Setup and Operation of the Security Advisory Board (SAB)

The data to be collected/generated in the framework of this activity fall under Tasks 8.2 and 9.2 (UNINOVA) and will be qualitative in nature. The data regarding the setup of the SAB include personal information of SAB members (e.g., name and surname, gender, organisation, position, email, country of placement, brief description of professional profile). They will be collected by the consortium partners, under the guidance of UNINOVA, following the process of retrieving information from online public sources (professional websites, etc). The data storage formats will include text documents, images and presentations (.docx,.ppt, .jpeg, .png, .pdf).

Data collected/generated through direct input methods will be stored in formats which allow the documentation of information from various files and documents in a single location. By doing so, it is possible to circulate raw data from transcripts, as well as text, images, and other objects from other files to one document file or multiple tabs of a single spreadsheet. Moreover, these formats can be immediately converted into open and machine-readable formats (e.g., .xml and .csv) boosting the interoperability and re-usability of the data produced in the framework of COLOSSUS.

2.2.2. Data collected/generated by the COLOSSUS system

Within the context of the implementation, testing and deployment of the COLOSSUS system, diverse datasets will be collected across its work packages (WPs). The following sub-sections outline the specific data types and their applications within the project (continuing the previous section's numbering).

10) USV data collection, processing and decision making

Within the COLOSSUS project, data collected and generated by the **Unmanned Surface Vessel (USV)** are essential for system development, testing, validation, performance assessment, and operational deployment in port and shipyard environments. At the **sensing and perception level**, the USV acquires environmental, navigational, and surveillance data through onboard awareness sensors. These include visual data (electro-optical and infrared imagery), sonar and underwater mapping data, LiDAR point clouds, Global Navigation Satellite System (GNSS) positioning data, and inertial measurements (IMU). These datasets support core functions such as autonomous navigation, obstacle avoidance, intrusion detection, underwater inspection, and three-dimensional reconstruction of port infrastructure. During testing and validation phases, the same data are also used to train, validate, and benchmark AI-based detection, perception, and swarm-coordination algorithms. At the **navigation, control, and propulsion level**, telemetry and system performance data are generated by the Pixhawk autopilot, VESC motor controllers, and PLC-based platform services. These datasets include heading and speed information, waypoint tracking logs, propulsion parameters, energy consumption metrics, and system health indicators. Such data are critical for evaluating autonomous navigation performance, verifying fail-safe mechanisms, assessing system robustness, and analysing resilience under stress-testing and degraded-operation conditions. At the **processing and intelligence layer**, higher-level datasets derived from

sensor fusion and AI algorithms are generated. These include object detection and tracking outputs, anomaly and event flags, classification results, mission execution logs, and decision-making traces. These data support the evaluation of algorithm accuracy, swarm coordination efficiency, response times to simulated threats, and overall system performance during laboratory experiments and operational trials. At the **communication and command-and-control layer**, data related to network performance and operator interaction are collected. These include communication latency, bandwidth usage, and link stability across multiple channels (e.g. 4G, UHF, satellite), as well as video streaming logs and operator interface interaction records. These datasets are used to validate communication resilience, assess real-time supervision and control capabilities, and evaluate the effectiveness of the human-in-the-loop decision-making model.

11) UAVs and UGVs data collection, processing and decision making

UAVs and UGVs collect visual and spatial data to support perception, navigation, and surveillance tasks in port environments. RGB and RGB-D data are acquired through onboard camera sensors on both platforms, capturing the surrounding environment and objects of interest such as personnel or infrastructure elements. These data streams are synchronised using timestamps and frame alignment and are represented using standard image and depth encodings, together with associated camera calibration information. In addition, UGVs collect three-dimensional point cloud data using 360° LiDAR sensors. These datasets capture structural and spatial characteristics of the environment and support mapping, localisation, navigation, and surveillance functions. Point cloud data are stored in efficient formats that allow the representation of multi-dimensional point attributes (e.g. spatial coordinates, intensity, and colour), enabling scalable processing of large and complex environments.

12) Secure interfaces for port security information exchange

The Secure Interface application server neither generates nor collects data. However, it facilitates encrypted data exchange between COLOSSUS subsystems and external entities. To this end, it may maintain separate databases for: (a) user access control and authorization management, (b) event logs, timestamps, and human approvals to ensure accountability, (c) mission parameters including position, attitude, speed, and trajectory logs, and (d) UAV/USV/UGV configuration management

13) Multi-modal data-efficient harmonisation

For this activity, the work focuses on the **use, processing, and exchange of data generated by other COLOSSUS system components**, rather than on the creation of new primary datasets. The data are utilised to support system integration, coordination, and decision-making processes within the USV-related workflows. This includes the analysis of perception outputs, mission status information, and coordination data to validate interoperability, assess system-level performance, and support operational testing, without resulting in the generation of additional standalone datasets.

14) Intelligent Swarm navigation for diverse UV types

This activity focuses on the processing and exchange of data generated by other COLOSSUS components, without creating new primary datasets. Data such as mission coordination messages, relative positioning and state information, task allocation data, and navigation constraints support cooperative behaviours including formation control, area coverage, collision avoidance, and dynamic re-tasking.

15) 3D reconstruction and modelling for underwater inspection

This activity uses sensor data generated by other COLOSSUS components to produce three-dimensional reconstructions of underwater port and maritime infrastructure. Sonar and visual data are processed to support inspection, anomaly detection, structural assessment, and monitoring of submerged assets. The resulting models are used for analysis, visualisation, and validation within the project and do not involve the generation or processing of personal data.

16) Integration lab testing and fine tuning

This activity involves the processing of technical data generated in other tasks, without creating new primary datasets. Data such as configuration parameters, system logs, performance metrics, interoperability test results, and error reports are used to validate integration, calibrate sensors and algorithms, refine component interfaces,

and improve overall system reliability. All data are non-personal and used exclusively for technical validation and optimisation.

17) COLOSSUS Web Platform

This activity focuses on the processing and visualisation of operational data generated by COLOSSUS system components, without generating new primary datasets. The web platform aggregates mission status information, alerts, system logs, performance indicators, and user interaction events to support monitoring, coordination, and human-in-the-loop decision-making. Data are used for system evaluation and demonstration purposes and are handled using appropriate access control, logging, and cybersecurity measures in line with project and regulatory requirements.

18) Pilot Preparation and Deployment

Each pilot operation must be planned well in advance to identify and mitigate any information that cannot be collected due to security restrictions. A **'Plan-Identify-Collect-Validate'** workflow shall be established for every use case to pre-define expected data and formats (e.g., .jpg, .mp4). This structured approach ensures that collected data is only released after a formal validation step: for instance, visual data (photos or videos) may be shared publicly provided that sensitive elements, such as Navy vessels, military personnel, or critical shipyard infrastructure, are blurred or removed. This methodology ensures project transparency and replicability while strictly maintaining operational security compliance.

19) Technical evaluation and market validation

Evaluation must be conducted under real-world shipyard conditions to ensure market readiness and scalability. Technical validation shall focus on operational KPIs that demonstrate value to the wider maritime sector: reduction in docking and undocking lead times, precision in underwater hull defect detection (UC#2), and automated response efficiency to perimeter breaches (UC#1). Success depends on verifying the Human-Machine Interface (HMI) with shipyard personnel to ensure the system is intuitive for non-technical operators. Furthermore, market validation must confirm system resilience against salt-water corrosion, tidal currents, and industrial signal interference, providing a proven 'blueprint' for the digital transformation of European naval and port infrastructures.

2.2.3. Data collected / generated from dissemination, communication and stakeholder engagement activities

The data derived by monitoring and assessment of the dissemination, communication activities of the project and by the stakeholder engagement with a view to measuring the impact of the relevant activities, fall under Tasks 6.1, 7.1 (Q-PLAN) and include:

20) Website analytics

These data will be collected/generated through periodic monitoring of the project's website statistics (online analytics, google analytics) with a view to measuring and assessing the performance and results of the project's website activity in terms of dissemination and communication. With that in mind, the data will be mostly quantitative in nature addressing key metrics (e.g., number of visitors, average page views per session, session duration, average time on page, country of visitors, etc.). All in all, the data will be stored in a spreadsheet (.xlsx or .csv) while at the same time the analysis of the results will be stored in a standard word document (.docx).

21) Social media account statistics

These data will be collected/generated through a periodic monitoring of the project's social media (i.e., LinkedIn, X, and YouTube) statistics with a view to measuring and assessing the performance and results of the project's social media activity in terms of dissemination and communication. With that in mind, the data will be mostly quantitative in nature addressing the metrics reached on each channel (e.g., number of followers, impressions on X, followers, likes, number of people reached through posts, etc.). All in all, the data will be stored in a spreadsheet (.xlsx) while at the same time the analysis of the results will be stored in a standard word document (.docx).

22) Data collected from project events

These data will be collected during the implementation of the project through: (i) the different events (e.g. co-creation workshops, physical and virtual events, etc.) organised by COLOSSUS (either alone or jointly with other projects or initiatives) consisting of the participants lists that will contain demographic information about the participants; and (ii) the participation of COLOSSUS partners in relevant third party events in order to reach out and engage stakeholders, thus collecting general information about the events attended and their outreach. Along these lines, these data will be collected to keep track of the results of events for stakeholder engagement and enable reporting on them. Moreover, these data will be updated every time a partner attends an event, or a partner organises an event. Finally, the data will be both quantitative and qualitative in nature and will be stored in a standard spreadsheet (.xlsx).

23) Newsletter subscription data

A subscription form hosted in the project's web site will aid the collection of these data in which any interested stakeholder can freely provide their contact details in a dedicated sign-up form so as to receive the most up-to-date news and outcomes of the project. A newsletter will be sent to subscribers once per 6 months. With that in mind, these data will be collected so as interested stakeholders can be informed about COLOSSUS' progress and upcoming events. The data will be comprised of a list of subscribers along with their personal information such as: (i) email address, (ii) first and last name, (iii) country, (iv) type of organisation, and (v) region. A copy of this contact list will be stored to MailChimp's server (http://mailchimp.com), which is used for e-mail campaigns and newsletters distribution. All personal information included in this contact list will be used and protected according to MailChimp's Privacy Policy. Subscribers can find details about how their personal information are managed on the respective privacy policy section of the COLOSSUS website.

24) Data collected from dissemination and communication activities

These data will be collected through the periodic monitoring of the project's miscellaneous (i) dissemination activities (e.g., participation in conferences, workshops, and external events, interviews) and (ii) communication activities (e.g., social media and blog posts, press releases). The data will be integrated in a single spreadsheet (.xlsx or .csv) designed to keep track of information related to such activities (e.g., dates, place, audience size for dissemination activities; target audience, social media URL, number of impressions for communication activities). The purpose of collecting these data is to assess the outreach of these activities during the implementation of the project. For this purpose, a template will be shared with all partners to log the activities they performed. The template will be provided online as well, so as the partners can directly update their input.

Q-PLAN is responsible for sending the necessary templates (.docx, .xlsx) to all partners, alongside with guidelines on how to fill them in, as well as for collecting input on an ad-hoc basis (i.e., each time a dissemination or stakeholder engagement action is performed). Q-PLAN is also responsible for preparing the necessary reports to evaluate the overall progress of dissemination and communication activities (measuring outcomes against pre-set Key Performance Indicators (KPIs)) throughout the lifespan of the project. The storage format of the data to be collected during the project's duration, includes .csv, .docx, .xlsx, .pdf, .pptx, .jpeg, and .png files.

2.3. Origin of data and re-use of pre-existing data

In the context of COLOSSUS, **new data** will be collected/generated by partners as well as external stakeholders participating in the activities of the project. With that in mind and apart from consortium partners, **external groups of stakeholders from which new data will originate include**:

- Port authorities' personnel, security personnel, coast guard personnel, that may serve as end-users of the COLOSSUS solutions.
- Knowledge, technology, and innovation solution providers (e.g., within academic institutions, non-university public research organisations, technology organisations, high-tech SMEs and large enterprises, etc.).
- AI and UV experts with scientific and industrial expertise.
- Relevant Initiatives (EU projects and working groups focusing on maritime border surveillance);

- Policy makers at regional, national and EU level (related to EU port resilience and maritime affairs)

Moreover, **pre-existing data (i.e. MAGIC Dataset)**[4] will be utilised within the context of COLOSSUS. Relevant data will be harvested from trusted selected online sources regarding port/shipyard security. Finally, consortium partners' internal knowledge, experience and expertise from their participation in other projects and initiatives will directly and indirectly support the implementation of activities throughout the project.

2.4. Expected size of data

COLOSSUS entails a series of activities aiming at setting the stage for and ultimately facilitating the demand-driven evidence-based development, piloting, evaluation, validation and fine-tuning of the COLOSSUS system and value propositions. With that in mind, the table that follows presents the different activities implemented during the course of the project in which data is collected/generated, the types and formats of the data as well as the expected size of the data.

Table 3 - Expected Size of Data

No	Name of Activity	Data	Type of Data	Data Form	Expected Data size
1	Market analysis	Desk research, semi-structured interviews and online survey outcomes	Interview transcripts, notes, text documents, survey outcomes	.docx, .xlsx, .csv	< 50 MB
2	Identification of regulatory, ethics and security standards requirements	Desk research and co-design outcomes	Notes, Text documents, presentations, guidelines	.docx, .pptx	< 50 MB
3	Co-design of the COLOSSUS requirements and system architecture	Co-design outcomes	Notes, text documents, presentations	.docx, .pptx	< 100 MB
4	Pilot Use Case Scenarios Planning	Co-design outcomes	Notes, text documents, presentations, photos	.docx, .pdf, .pptx, .jpeg, .png	< 100 MB
5	Replication guidelines	Best practice documentation, scalability assessment data, and market access guidelines for SMEs	Text documents, spreadsheets	.docx/ .xlsx	< 5 MB
6	Policy recommendations	Outputs underpinning policy recommendations	Text documents	.docx	< 50 MB
7	Business development	Desk research, semi-structured interviews, interaction with industry actors	Interview transcripts, notes, text documents	.docx	< 50 MB

[4] Ma, W., Zhong, H., Wang, L., Jiang, L. and Abdel-Aty, M., 2022. MAGIC dataset: Multiple conditions unmanned aerial vehicle group-based high-fidelity comprehensive vehicle trajectory dataset. *Transportation research record*, *2676*(5), pp.793-805.

8	Collaboration and synergies with relevant projects and initiatives	Collaboration- and synergy-related data	Notes, spreadsheets, images, presentations	.docx, .xlsx, .pdf, .pptx, .jpeg, .png	< 200 MB
9	Setup and Operation of the SAB	SAB membership and activities	Notes, spreadsheets, images	.docx, .xlsx, .pdf, .jpeg	< 100 MB
10	USV data collection, processing and decision making	Integrated multi-sensor data to support the decision-making process.	Time Series	JSON	N/A
11	UAVs and UGVs data collection, processing and decision making	RGB and RGB-D data collected from camera sensors	Images	.png	> 5 GB
12		Point cloud data from vehicle sensors (e.g., LiDAR) or estimation form depth data.	Point Cloud	.laz, .ply, .pcd	1-10 GB
13	Secure interfaces for port security information exchange	Access control and authorization data, event logs, mission parameters, trajectory logs, UAV/USV/UGV configuration	Logs, events, structured data	Yaml, json, xml	<1GB
14	3D reconstruction and modelling for underwater inspection	Point cloud representing the 3D reconstruction model, with each point associated to its 2D image counterpart (if reconstructing from 2D imagery).	Dense Point Cloud	.ply	5-20 GB
15	Integration lab testing and fine tuning	Integrated multi-sensor data to support the decision-making process.	Time Series	JSON	N/A
16	Lab stress-testing and technology validation	Integrated multi-sensor data to support the decision-making process.	Time Series	JSON	N/A
17	COLOSSUS Web Platform	Integrated multi-sensor data to support the decision-making process.	Time Series SQL	JSON	N/A
18	Pilot Preparation and Deployment	Operational Pilot Planning & Mitigation Logs	Qualitative and Quantitative	To be determined during pre-pilot planning (T1.4)	N/A

19	Technical evaluation and market validation	Technical Performance & Market Validation Metrics	Quantitative (Performance KPIs)	.xlsx, .csv, .pdf	N/A
20	Monitoring and assessment of the dissemination, communication and stakeholder engagement activities	Website analytics	Quantitative data (e.g., No of unique visitors)	.csv, .docx, .xlsx, .pdf, .pptx, .jpeg, .png	< 20 MB
21		Social media account statistics	Quantitative data (e.g., Followers, posts/tweets as well as click stats, engagement rates, impressions etc.)	.csv, .xlsx	< 20 MB
22		Data collected from project events	Spreadsheets,	.xlsx, .jpeg, .png	< 1 MB
23		Newsletter subscription data	Spreadsheets	.csv, .xlsx,	< 1 MB
24		Data collected from dissemination and communication activities	Spreadsheets	.csv, .xlsx, .pdf, .jpeg, .png, .docx	< 500 MB

2.5. Data Utility

The stakeholders that may find meaningful utility for the data to be collected/generated by the project (both within as well as outside of COLOSSUS' consortium) along with the benefits that could arise for them by utilizing these data, are concisely presented in the table that follows.

Table 4 - Data Utility

Stakeholder Group	Data Utility
Technology providers & Commercial/Industry stakeholders	The data to be collected/generated from COLOSSUS' activities can be useful to technology or industry stakeholders who are interested in developing their own applications in the field of port/shipyard resilience (incl. security, operations, and predictive maintenance). The datasets, especially those derived from use cases, could be used to provide evidence for the benefits of integrated port resilience solutions. Moreover, the market research data and insights gathered from ecosystem stakeholders could be useful to understand prevailing market trends at this moment in time.
Civil security adopters / Security practitioners	Civil security adopters and/or security practitioners can make use of COLOSSUS' data to understand how their operations could benefit from utilising solutions similar to the one generated by the project. Operational testing data would be well suited to giving them insights into how such solutions perform under different conditions.
Academia	COLOSSUS will provide the academic community and researchers with relevant datasets that could be utilised for further analysis and research in order to address a wide range of technical challenges in many fields, such as unmanned vehicles, swarm navigation, object recognition, 3d modelling, AI-based approaches, and in overall how to approach security projects.
Policymakers and regulators	COLOSSUS is expected to generate data that will help policymakers to better understand how EU policies can not only better secure critical infrastructure

Stakeholder Group	Data Utility
	such as ports and shipyards but also facilitate access to cutting edge innovations supporting mor cost-effective technology adoption. Regulators may use COLOSSUS' data to update regulatory frameworks for the technologies studied and developed by the project.

While **Chapter 2** outlines all activities and data generated, the specific algorithms, software components, trained models, and designs will also be managed as "Other research outputs" as detailed in **Chapter 4.**

3. FAIR data

The guidelines on Data Management Plan[5] of the Commission emphasise the importance of making the data produced by projects funded under Horizon Europe **Findable, Accessible, Interoperable as well as Reusable (FAIR)**, with a view to ensuring their sound management. This means using standards and metadata to make data discoverable, specifying data sharing procedures and which data will be open, allowing data exchange via open repositories as well as facilitating the reusability of the data. With that in mind, the following sections of the DMP lay out the methodology followed in the framework of COLOSSUS with respect to making data findable, accessible and interoperable as well as ensuring their preservation and open access, with a view to increasing their re-use.

3.1. Making data findable, including provisions for metadata

3.1.1. Data discoverability and identification mechanisms

COLOSSUS places special emphasis on enhancing the discoverability of the data collected/generated or re-used during the course of its activities. **Open data produced during the implementation of the project will be locatable by means of a standard identification mechanism.** Indeed, COLOSSUS will be able to assign globally resolvable **Persistent Identifiers (PIDs)** on any open data (more information on open data as well as the respective repositories we plan on employing in the context of the project are provided in section 3.2). An identifier is a unique identification code that is applied to a dataset, so that it can be unambiguously referenced. For example, a catalogue number is an identifier for a particular specimen, and an ISBN code is an identifier for a particular book. PIDs are simply maintainable identifiers that allow for permanent reference to a digital object. In other words, PIDs are a way of giving digital resources, such as documents, images and data records, a unique and persistent reference number.

At the same time, data that are not open will be deposited in a searchable resource (i.e., the cloud web storage service of the project) and well-tailored identification mechanisms will be utilized as well, in the form of standard naming conventions that will safeguard their consistency and make them easily locatable for partners within the frame of the project. Along these lines, the following subsection provides further analysis on naming conventions and versioning.

3.1.2. Naming conventions and versioning

Following a consistent set of naming conventions in the development of the project's data files can greatly enhance their searchability. With that in mind, COLOSSUS creates consistent data file names that provide clues

[5] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

to their content, status and versioning, while also increasing their discoverability. In doing so, project partners as well as interested stakeholders can easily identify a file as well as classify and sort them.

According to the UK Data Archive (*UK Data Service, 2017b*), a best practice in naming convention is to create brief yet meaningful names for data files, that facilitate classification. The naming convention should avoid the utilisation of spaces, dots and special characters (such as & or !), whereas the use of underscores is endorsed, to separate elements in the data file name and make them understandable. At the same time, versioning should be a part of a naming convention to clearly identify the changes and edits in a file.

To facilitate the reference of the datasets that will be produced during its implementation, COLOSSUS employs a **standard naming convention that integrates versioning and takes into account the possibility of creating multiple datasets** during an activity that entails data collection/generation. Indeed, COLOSSUS' naming convention considers this issue and addresses it by employing a unique element that captures the number of datasets that are produced under the same activity.

In particular, the **naming convention employed by the project is described below**.

[Name of project] _ [Name of Study] _ [Number of dataset] _ [Version number] _ [Issue Date]

- **Name of project:** COLOSSUS
- **Name of Study:** A short version of the name of the activity for which the dataset is created.
- **Number of dataset:** An indication of the number assigned to the dataset.
- **Version number:** The versioning number of a dataset.
- **Issue Date:** The date on which the latest version of the dataset was modified (YYYY.MM.DD.).

With the above in mind, some **indicative examples** to showcase the naming structure that will be applied in the context of COLOSSUS are provided below:

- **COLOSSUS_InterviewsUsercenteredRequirements_Dataset1_v1_2026.03.31 -** The first dataset generated from the interviews conducted with key public and private stakeholders from the consortium ecosystem, as well as external entities, to identify user-centred requirements, needs and challenges. This is the first version of the dataset that was last modified on the 31st of March 2026 (31/03/2025);
- **COLOSSUS_UC#1_Dataset2_v2_2027.07.31 –** The second dataset created in the process of preparing and conducting the 1st Use Case of COLOSSUS. The last modification of this dataset, which in this case produced the second version of the dataset, was on the 31st of July 2027 (31/07/2027).

Versioning of information makes a revision of datasets uniquely identifiable and can be used to determine whether and how data changed over time and to define specifically which version the creators/editors are working with. Moreover, effective data versioning enables understanding if a newer version of a dataset is available and which are the changes between the different versions allowing for comparisons and preventing confusion. In this context, **a clear version number indicator is used in the naming convention** of every data file produced during COLOSSUS in order to facilitate the identification of different versions.

3.1.3. Metadata allowing discovery

In addition to consistent naming conventions and versioning, the project also follows a metadata-driven approach so as to allow discovery and further increase the searchability of the data, while also facilitating its understanding and re-use. Metadata is defined as "data about data" or "information about information"[6]. It is usually structured textual information that describes the creation, content, or context of a digital resource – be it a single file, part of a single file, or a collection of many files. Metadata is the glue, which links information and

[6] Huxley, L., & Jacobs, N. (2004). Online information services in the Social Sciences. Oxford: Chandos.

data across the world wide web. It is the tool that helps people to discover, manage, describe, preserve and build relationships with and between digital resources [7].

In particular, three distinct types of metadata exist[8], as presented below:

- **Descriptive metadata** is used to identify and describe collections and related information resources. Descriptive metadata at the local level helps with searching and retrieving. In an online environment, descriptive metadata helps to discover resources. Most of the times, it includes information such as the title, author, date, description, identifier, etc.
- **Administrative metadata** is used to facilitate the management of information resources. It is helpful for both short-term and long-term management and processing of data. This is information that will not usually be relevant to the public but will be essential for staff to manage collections internally. Such metadata may be location information, acquisition information, etc.
- **Structural metadata** enables navigation and presentation of electronic resources. It documents how the components of an item are organized. Examples of structural metadata could be the way in which pages are ordered to form chapters of a book, a photograph that is included in a manuscript or a scrapbook or the JPEG and TIF files that were created from the original photograph negative, linked together.

With that in mind, **data produced/used during COLOSSUS is discoverable with metadata** suitable to its content and format. The project employs **metadata standards** to produce rich and consistent metadata with a view to supporting the long-term discovery, use and integrity of its data. More details on the metadata standards adopted by COLOSSUS are provided on the following subsection.

3.1.4. Standards for metadata creation

COLOSSUS employs standards for creating metadata for data collected/generated by the project, with a view to describing it with **rich metadata** and thus improving their discoverability and searchability. In result, effective searching, improved digital curation and easy sharing will be realized. In addition, the metadata standards applied enable the integration of metadata from a variety of sources into other technical systems.

With that in mind, for **COLOSSUS' openly available data** the metadata standards provided by **Zenodo and/or GitHub will be used**. Zenodo (*https://zenodo.org/*) is an open repository developed under the European OpenAIRE programme and operated by CERN (European Organization for Nuclear Research). The repository along with its metadata standards have been adopted and are being used by numerus research communities, enabling them to deposit research papers, datasets, software, reports as well as other research outputs. Along these lines, Zenodo creates metadata to accompany the datasets that are uploaded to the repository, extending their reach to a wider audience of interested stakeholders. These metadata can be exported in several standard formats, including open and machine-readable ones (such as MARCXML, Dublin Core, and DataCite Metadata Schema), following the guidelines of OpenAIRE and are stored by Zenodo in JSON-format according to a defined JSON schema[9]. Similarly, GitHub (*https://github.com/*) is a cloud-based platform where you can store, share, and work together with others to write code. Github exposes useful metadata for each (code) repository as environment variable which can be called / used in a workflow[10].

Project **data not open (closed data), will also be annotated with open and machine-readable metadata** following the **Dublin Core Metadata standard**. The Dublin Core Metadata element set (certified with the ISO Standard 15836) is a standard which can be easily understood and implemented and as such, is one of the best-known metadata standards. It was originally developed as a core set of elements for describing the content of web pages and enabling their search and retrieval. Among the reasons for selecting this standard is also the fact that **Zenodo is compatible with Dublin Core metadata formats** and thus any initially closed data, that

[7] Foulonneau, M., & Riley, J. (2008). Metadata for digital resources: Implementation, systems design and interoperability. Oxford: Chandos.
[8] Caplan, P. (2003). Metadata fundamentals for all librarians. Chicago: American Library Association.
[9] For more information on the JSON format and the JSON schema visit the following website: http://json-schema.org/
[10] https://github.com/marketplace/actions/github-repository-metadata

may become open at a later stage (e.g., due to a change in the consortium's policy), will not lose its metadata. With that said, the Dublin Core metadata standard is a simple yet effective set for creating rich metadata that will describe a wide range of resources. The fifteenth element "Dublin Core" described in this standard is part of a larger set of metadata vocabularies and technical specifications maintained by the Dublin Core Metadata Initiative (DCMI)[11]. The full set of vocabularies also includes sets of resource classes, vocabulary encoding schemes, and syntax encoding schemes. **An online metadata generator will be used** to produce the different metadata elements required (*dublincoregenerator.com*).

3.1.5. Search keywords included in the metadata

The project's data will be provided with search keywords with a view to optimizing its findability as well as its ultimate re-use by interested stakeholders during its entire lifetime. With that in mind, the metadata standards employed by COLOSSUS provide opportunities for tagging the data collected/generated and its content with keywords. In general, keywords are a subset of metadata and include words and phrases used to name data. In the context of COLOSSUS, keywords are used to add valuable information to the data collected/generated as well as to facilitate the description and interpretation of its content and value.

Along these lines, the project's strategy on keywords is underpinned by the following principles:

- The who, the what, the when, the where, and the why should be covered;
- Consistency among the different keyword tags needs to be ensured;
- Relevant, understandable and clear keywording ought to be sought.

In general, the keywords will comprise terms related to port resilience, security, operations, UVs, AI, etc. The keywords will accurately reflect the content of the datasets and avoid words used only once or twice within them.

3.1.6. Offering metadata that can be harvested and indexed

We know that the wild diversity of the metadata accompanying open data across the plethora of online repositories (e.g., disciplinary archives, institutional repositories, open access journals) can serve as barriers for their findability and sharing amongst different research communities. This is why in the context of COLOSSUS we have aligned our metadata creating approach with the **Open Archives Initiative (OAI)**, which promotes the use of a standard protocol for metadata harvesting, designed for better sharing and retrieval of data residing in distributed repositories. This protocol, namely the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)[12], promotes interoperability standards that facilitate efficient dissemination of data amongst diverse communities[13].

All structured **metadata linked to the project's open data will be offered in a way that can be exported and harvested** via the OAI-PMH thanks to the standards we adopt for metadata creation (see section 3.1.4). The same standards will also help us produce **metadata that facilitate indexing**. For instance, the use of the Dublin Core Metadata Standard (as further elaborated in section 3.3) provides a vocabulary of concepts with definitions in open-machine readable formats that enable easier indexing of metadata. Along these lines, there are several tools[14] which implement the Archives Initiative Protocol for Metadata Harvesting, such as **Arc source**, **EnhancedOAIServer** and **eprints.org**, and can be used for harvesting our data by different repositories.

COLOSSUS' openly available data will be uploaded in Zenodo, which is in line with FAIR principles, including the "To be Findable" principle. Metadata of each record uploaded in Zenodo is indexed and searchable directly in Zenodo's search engine immediately after publishing. Metadata of each record is sent to DataCite servers during DOI registration and indexed there.

[11] Retrieved from: https://www.dublincore.org/
[12] Retrieved from: https://www.openarchives.org/pmh/
[13] Corrado, E.M. (2005) 'The importance of open access, open source, and open standards for libraries', Issues in Science and Technology Librarianship.
[14] For more information about the tools implementing the OAI-PMH: https://www.openarchives.org/pmh/tools/

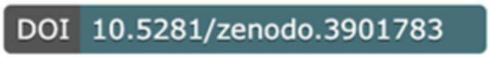
3.2. Making data accessible

3.2.1. Repositories

The **data produced by COLOSSUS and deemed open for sharing and re-use, will be deposited to and securely stored by Zenodo** (*www.zenodo.org*), **which constitutes an open data repository** and has been specifically selected to enable access to the project's open data free of charge. In fact, Zenodo builds and operates a simple service that enables researchers, scientists, EU projects and institutions, among others, to share and showcase research results (including data and publications) that are not part of the existing institutional or subject-based repositories of the research communities. It accepts any file format, promotes peer-reviewed openly accessible research, allows the creation of own collections and it is available free of charge both for COLOSSUS to upload and share data as well as for other stakeholders to explore, download and re-use these data.

Moreover, as a digital repository, Zenodo registers **Digital Object Identifiers (DOIs)** for all submitted data through DataCite[15], which is the leading global non-profit organisation that provides PIDs (and specifically DOIs) for research data and preserves these submissions using the safe and trusted foundation of CERN's data centre, alongside the biggest scientific dataset in the world, the LHC's 100PB Big Data store[16]. This means that the data preserved in Zenodo will be accessible for years to come, and the DOIs will function as perpetual links to the resources. DOIs remain valuable since they are future-proofed against Uniform Resource Locator (URL) or even protocol changes, through resolvers (such as DOI[17]). With that in mind, an example of a DOI retrieved from this open repository follows the structure illustrated in Figure 1.

Figure 1: Typical DOI created by Zenodo.



Moreover, all **non-sensitive software components and open-source code will be deposited and securely stored in GitHub (www.github.com),** hosted by UNINOVA. GitHub has been selected as an appropriate repository because it is the world's most widely used platform for open-source software development, offering robust version control, collaboration tools, and long-term public accessibility free of charge. It supports distributed version control through Git, issue tracking, collaborative review workflows, continuous integration, and automated testing[18]. This enables COLOSSUS partners to maintain transparent and reproducible development practices while facilitating contributions from external developers where appropriate. GitHub accepts all standard code and text-based file formats and allows the creation of dedicated organisation-level repositories and project-specific structures, ensuring clear organisation and discoverability of the COLOSSUS codebase[19].

All public repositories on GitHub are accessible free of charge to external stakeholders, who can view, clone, download, fork, and re-use the COLOSSUS code in accordance with its open-source licence[20]. The platform supports machine-readable metadata files (e.g., README, LICENSE, and CITATION.cff)[21]. GitHub also integrates seamlessly with Zenodo to provide persistent DOIs for software releases. By linking the COLOSSUS GitHub repositories with Zenodo, each versioned release of the project's software can be assigned a DOI through DataCite, ensuring that the software is permanently citable, archived, and accessible even if repository structures change in the future[22].

[15] For more information on DataCite: https://www.datacite.org/
[16] Retrieved from: https://www.software.ac.uk/tags/zenodo
[17] Retrieved from: http://dx.doi.org/
[18] GitHub. "About GitHub." Retrieved from https://docs.github.com/en/get-started/using-github/about-github
[19] GitHub. "Creating an organization. Retrieved from https://docs.github.com/en/organizations/collaborating-with-groups-in-organizations/about-organizations
[20] GitHub. "Working with forks." Retrieved from https://docs.github.com/en/pull-requests/collaborating-with-pull-requests/working-with-forks
[21] GitHub. "CITATION files." Retrieved from https://docs.github.com/en/repositories/managing-your-repositorys-settings-and-features/customizing-your-repository/about-citation-files
[22] Zenodo. "GitHub Integration." Retrieved from https://help.zenodo.org/features/github

Furthermore, any relevant, open-source AI assets will be stored on the AI on Demand (AIoD) platform (www.aiod.eu). AIoD is Europe's central hub for AI research, innovation, and deployment assets and is appropriate because it provides broad access to AI assets and supports the FAIR principles of findability and re-usability across distributed sources[23].The platform functions as a **unified metadata catalogue** that aggregates AI assets from across the research and innovation ecosystem, including datasets, models, software, tools, documentation, publications, and educational resources. It supports **search, browsing, and integration via web interfaces, software development kits (SDKs), and APIs**, enabling both human users and automated tools to discover and access resources relevant to COLOSSUS[24].

3.2.2. Data

3.2.2.1 Openly available and closed data

COLOSSUS, in line with FAIR principles of data management in the context of Horizon Europe, adopts the good practice of making data as open as possible and as closed as necessary. This calls for partners to disseminate their data that have the potential to offer long-term value to external stakeholders and do not harm the confidentiality and privacy of the stakeholders that contributed to the collection/generation of these data, maximising the beneficial impact of COLOSSUS.

Only anonymised and aggregated data will be made open (publicly available) to ensure that data subjects cannot be identified in any reports, publications and/or datasets resulting from the project. The relevant project partner in each case will **undertake all the necessary anonymisation procedures** to anonymise the data in such a way that the data subject is no longer identifiable (more details on data management responsibilities are provided in **Section 5.2**), in compliance with applicable European and National Regulations for personal data protection.

To this end, it is important to keep in mind that during the process of data anonymisation, data identifiers need to be removed, generalised, aggregated or distorted. Moreover, **anonymisation is different than pseudonymisation**, which falls under a distinct category in the GDPR - anonymisation theoretically fully and irreversibly prevents any backtrace of the subject's data, while pseudonymisation allows for the data subject to be re-identified with additional information. Along these lines, the table which follows provides a **list of good practices** for the anonymisation of quantitative and qualitative data derived from the tour guide on data management of the Consortium of European Social Science Data Archives (CESSDA).

Table 5: Good practices for data anonymisation

Type of data	Good practices
Quantitative data	• Remove or aggregate variables or reduce the precision or detailed textual meaning of a variable. • Aggregate or reduce the precision of a variable such as age or place of residence. As a general rule, report the lowest level of geo-referencing that will not potentially breach respondent confidentiality. • Generalise the meaning of a detailed text variable by replacing potentially disclosive free-text responses with more general text. • Restrict the upper or lower ranges of a continuous variable to hide outliers if the values for certain individuals are unusual or atypical within the wider group researched.
Qualitative data	• Use pseudonyms or generic descriptors to edit identifying information, rather than blanking-out that information. • Plan anonymisation at the time of transcription or initial write-up, (longitudinal studies may be an exception if relationships between waves of interviews need special attention for harmonised editing).

[23] AI-on-Demand Platform. "About" Retrieved from https://aiod.eu/get-started/
[24] AI-on-Demand Platform. "AI Catalogue" Retrieved from https://aiod.eu/metadata-catalogue-rest-api/

Type of data	Good practices
	• Use pseudonyms or replacements that are consistent within the research team and throughout the project. For example, using the same pseudonyms in publications and follow-up research. • Use 'search and replace' techniques carefully so that unintended changes are not made, and misspelt words are not missed. • Identify replacements in text clearly, for example with [brackets] or using XML tags such as <seg>word to be anonymised</seg>. • Create an anonymisation log (also known as a de-anonymisation key) of all replacements, aggregations or removals made and store such a log securely and separately from the anonymised data files.

Source: Tour guide on data management of the CESSDA[25]

With that in mind, the following table presents the data collected/generated during the course of the project that will be made openly available. In case certain data cannot be shared (or need to be shared under restrictions), a justification for that choice is provided.

Table 6: Data availability

#	Data	Availability	Notes
1	Market analysis	Open & Closed	Data collected will be available only to the COLOSSUS consortium and the European Commission. In case there is a need to share information for dissemination and communication purposes, any sensitive information and identifiers related to the product development will be removed.
2	Identification of regulatory, ethics and security standards requirements	Open & Closed	Data collected will be available only to the COLOSSUS consortium and the EC. In case there is a need to share information for dissemination and communication purposes, any sensitive information and identifiers related to the product development will be removed.
3	Co-design of the COLOSSUS requirements and system architecture	Closed	Data collected will be available only to the COLOSSUS consortium and the EC.
4	Pilot Use Case Scenarios Planning	Closed	Data collected will be available only to the COLOSSUS consortium and the EC. In case there is a need to share information for dissemination and communication purposes, any sensitive information and identifiers related to the product development will be removed.
5	Replication guidelines	Open	-
6	Policy recommendations	Open	-
7	Business development	Closed	Data collected will be available only to the COLOSSUS consortium and the EC. In case there is a need to share information for dissemination and communication purposes, any sensitive information and identifiers related to the product development will be removed.

[25] Retrieved from: https://www.cessda.eu/Research-Infrastructure/Training/Expert-Tour-Guide-on-Data-Management/5.-Protect/Anonymisation

#	Data	Availability	Notes
8	Collaboration and synergies with relevant projects and initiatives	Open	-
9	Setup and Operation of the SAB	Open & Closed	The composition of the SAB will be publicly available through the project's website. All other data collected by the interaction with the SAB will be available only to the COLOSSUS consortium and the EC.
10	USV data collection, processing and decision making	Closed	Data collected will be available only to the COLOSSUS consortium and the EC.
11	UAVs and UGVs data collection, processing and decision making	Closed	Data collected will be available only to the COLOSSUS consortium and the EC.
12	Secure interfaces for port security information exchange	Closed	Data collected will be available only to the COLOSSUS consortium and the EC.
13	Intelligent Swarm navigation for diverse UV types	Closed	Data collected will be available only to the COLOSSUS consortium and the EC.
14	3D reconstruction and modelling for underwater inspection	Closed	Data collected will be available only to the COLOSSUS consortium and the EC.
15	COLOSSUS Web Platform	Closed	Data collected will be available only to the COLOSSUS consortium and the EC.
16	Pilot Preparation and Deployment	Open & Closed	Restricted / Validation Required. Planning logs and mitigation protocols will be available to the consortium. Public release is subject to the 'Plan-Identify-Collect-Validate' workflow to protect shipyard security.
17	Technical evaluation and market validation	Open	Public (Anonymized/Sanitized). Technical KPIs and performance results will be made available to ensure project replicability. Raw data containing sensitive infrastructure details will remain restricted.
18	Website analytics	Open & Closed	Website analytics will be available only to the COLOSSUS consortium and the EC. In case statistics are shared, data will be aggregated and anonymised before being made openly available, while personal data will be treated as expected by the GDPR.
19	Social media account statistics	Open & Closed	Social media analytics will be available only to the COLOSSUS consortium and the EC. In case statistics are shared, data will be aggregated and anonymised before being made openly available, while personal data will be treated as expected by the GDPR.
20	Data collected from project events	Open & Closed	Data collected from project events (e.g., Demo Days) will be available only to the COLOSSUS consortium and the EC. In case statistics are shared, data will be aggregated and anonymised before being made openly available, while personal data will be treated as expected by the GDPR.

#	Data	Availability	Notes
21	Newsletter subscription data	Closed	Data from newsletter subscriptions will remain closed as it contains personal information and is useful only for internal reporting purposes.
22	Data collected from dissemination and communication activities	Open & Closed	Data collected from dissemination and communication actions will be available only to the COLOSSUS consortium and the EC. In case there is a need to share information for dissemination and communication purposes, any personal information will be anonymised before being made openly available, while personal data will be treated as expected by the GDPR.

It is important to note that all personal data collected/generated will be considered as **closed data** prior to their anonymisation and aggregation to safeguard the confidentiality of the data subjects.

3.2.2.2 Data accessibility and availability

Public access to the open data will be made available and free of charge through Zenodo, which will automatically link to OpenAIRE, GitHub, and AIoD (depending on the type of data). The data will be fully accessible thanks to the included metadata and the search facility available on Zenodo. At the same time, closed data are intended to be stored and shared amongst authorised members of the consortium through cloud storage and file sharing providers which constitute structures that maintain and manage data and make these data accessible over a network, usually the internet (i.e., MS Teams). Before starting using these cloud services from providers situated both inside and outside the EEA, we have ensured that they comply with the relevant GDPR requirements.

The following table presents where data will be made accessible in the context of COLOSSUS.

Table 7: Data accessibility

#	Data	Accessibility notes
1	Market analysis	COLOSSUS website and Zenodo
2	Identification of regulatory, ethics and security standards requirements	COLOSSUS website and Zenodo
3	Co-design of the COLOSSUS requirements and system architecture	N/A - Closed data
4	Pilot Use Case Scenarios Planning	N/A - Closed data
5	Replication guidelines	COLOSSUS website and Zenodo
6	Policy recommendations	COLOSSUS website and Zenodo
7	Business development	N/A - Closed data
8	Collaboration and synergies with relevant projects and initiatives	COLOSSUS website and Zenodo
9	Setup and Operation of the SAB	COLOSSUS website
10	USV data collection, processing and decision making	N/A - Closed data

#	Data	Accessibility notes
14	Intelligent Swarm navigation for diverse UV types	N/A - Closed data
16	Integration lab testing and fine tuning	N/A - Closed data
17	Lab stress-testing and technology validation	N/A - Closed data
18	COLOSSUS Web Platform	N/A - Closed data
19	Pilot Preparation and Deployment	Restricted (Security). Risk mitigation during planning to prevent sensitive leaks.
20	Technical evaluation and market validation	OPSEC / Privacy. Data sanitized to remove military/personal identifiers before release.
21	Website analytics	COLOSSUS website
22	Social media account statistics	COLOSSUS website
23	Data collected from project events	COLOSSUS website
24	Newsletter subscription data	N/A - Closed data
25	Data collected from dissemination and communication activities	COLOSSUS website

3.2.2.3 Restrictions on use

By utilising Zenodo for sharing the project's openly available data, COLOSSUS can apply **different levels of accessibility** for these data taking into account any relevant issues (such as ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related, etc.).

More specifically, **Zenodo offers the following levels of data accessibility**:

- **Open access**: Data remains available for re-use. Nevertheless, the level in which these data can be re-used is determined also by their accompanied licence for re-use (see subsection 3.4.3);
- **Embargoed status**: Access to the data will be restricted until the end of the embargo period, at which time, the content will automatically become publicly available;
- **Restricted access**: The data will not be made publicly available and sharing will be made possible only by the approval of the project partner that have the responsibility of the data;
- **Closed access**: The data are protected against unauthorized access at all levels and only members of the consortium have the right to access it.

The same principles apply for GitHub (public/private repositories) and AIoD (where relevant), with differing sharing and access options.

Project partners will mainly use the open access level to disseminate the project's data amongst the interested stakeholders. Data that will not be available for re-use will be accessible only by authorised partners of COLOSSUS' consortium and/or authorised personnel from the funding authority of the project.

Moreover, **COLOSSUS will ensure open access to all peer-reviewed scientific publications** that may be produced in the framework of the project. In particular, according to the Grant Agreement, COLOSSUS will:

- at the latest at the time of publication, a machine-readable electronic copy of the published version or the final peer-reviewed manuscript accepted for publication, is deposited in a trusted repository for scientific publications;

- immediate open access is provided to the deposited publication via the repository, under the latest available version of the Creative Commons Attribution International Public Licence (CC BY) or a licence with equivalent rights; for monographs and other long-text formats, the licence may exclude commercial uses and derivative works (e.g., CC BY-NC, CC BY-ND);
- information is given via the repository about any research output or any other tools and instruments needed to validate the conclusions of the scientific publication.

Beneficiaries (or authors) must retain sufficient intellectual property rights to comply with the open access requirements.

3.2.2.4 Identity ascertainment and data access committee

The identity of stakeholders who want to access the open data on Zenodo, GitHub or AIoD, is not necessary to be ascertained, as the uploaded data are publicly open and no authorization is needed. On the other hand, closed for the public data will be available only to authorized consortium partners through dedicated mechanisms provided by the cloud storage service employed by the respective partners in order to deposit the data.

The need for a data access committee to evaluate or approve access requests to personal data, is not foreseen because only authorized partners will have access to the project's closed data, accessible only by using their credentials (username/password), and no third-party will re-use them for their benefit.

3.2.3. Metadata

3.2.3.1 Availability and licences

Metadata of deposited publications generated in the context of COLOSSUS will be **open under a Creative Common Public Domain Dedication (CC 0)** or equivalent, in line with the FAIR principles for data management adopted by the project (in particular machine-actionable). Such **metadata will provide information, at least, about the following**:

- The publication at hand (author(s), title, date of publication, publication venue);
- Reference to the Horizon Europe funding;
- The name of the project, including its acronym and Grant Agreement number;
- Any particular licensing terms which may apply (depending on the chosen license);
- Persistent identifiers that have been attributed to the publication;
- Authors involved in the action, their organisations and the project itself.

Where applicable, the metadata will also include persistent identifiers for any research output or any other tools and instruments needed to validate the conclusions of the publication. The metadata will be available through Zenodo, GitHub or AIoD. It is quite unlikely that Zenodo, GitHub or AIoD will terminate their operation and stop providing their services, but in such a case, all data, metadata, code and documentation uploaded will be transferred and hosted to other suitable repositories without undue delay. In this respect, it is important to note that, since all of COLOSSUS' openly available data will make use of PIDs (i.e., DOIs), which is further elaborated in subsection 3.1.1, the links to the data will not be affected. In parallel, the project's data that will not be openly available for sharing will be deposited, together with their accompanying metadata, code and documentation (if necessary), to the cloud web storage service employed by the project.

3.2.3.2 Methods, Software tools and documentation to access the data

COLOSSUS emphasises the accessibility of the data collected/generated during the project. With that in mind, no specialised method, software tool and/or documentation is expected to be needed at the moment, in order to access the data. Stakeholders will have the ability to access the data by simply using their web browser (e.g., Mozilla, Google Chrome, Internet Explorer, Safari, etc.) through their computers (either desktop or laptop), smartphones and/or tablets.

More specifically, through the COLOSSUS website clear links to the selected repositories will allow direct navigation to the publicly available datasets in all three online repositories (i.e., Zenodo, GitHub, and AIoD). In addition, users can access their websites and utilise the search engine of each repository to search for COLOSSUS' data. By typing the name of the project (or any other relevant keyword connected to the COLOSSUS data) the search engine will direct the user to the project's data, ready to be explored and re-used. Moreover, since the data will be available in open formats, we will be ensuring that they can appropriately be read by a range of different software that are widely and freely accessible to all potential users of the data.

Closed data will only be accessed by authorised project partners through the usage of a cloud storage service. Again, no specialised method, software tool and/or documentation is needed to this end.

As it was further elaborated in subsection 3.2.2, if Zenodo, GitHub or AIoD terminate their operation and stops providing its services, in such a case all data, metadata, code and documentation uploaded will be transferred and hosted to other suitable repositories without undue delay.

Along these lines, this section has provided the methodology applied in the frame of COLOSSUS to ensure that its data are as openly accessible as possible by any stakeholder that may find them interesting for re-use. In this context, COLOSSUS also focuses on providing metadata standards and appropriate metadata vocabularies to increase data interoperability. The following section provides further details in this respect.

3.3. Making data interoperable

Data interoperability refers to the ability of systems and services that create, exchange and use data to have clear, shared expectations for the contents, context and meaning of that data[26]. With that in mind, COLOSSUS has adopted in its data management methodology the use of metadata vocabularies, standards and methods that will increase the interoperability of the data collected/generated through its activities.

More specifically, **the interoperability of the data that will not be publicly shared will be facilitated by the use of the Dublin Core Metadata standard.** This standard is a small "metadata element set", which accounts for issues that must be resolved in order to ensure that data meet traditional standards for quality and consistency, while remaining broadly interoperable with other data sources in the linked data environment. The fifteen elements of the standard provide a vocabulary of concepts with natural-language definitions (e.g., title, creator, author, etc.) that are instantly converted into open machine-readable formats (such as XML, HTML, etc.), enabling machine-processability. Each element is optional and may be repeated, while the standard itself offers ways exist for refining them, encouraging the use of encoding and vocabulary schemes. The vocabulary of the Dublin Core Metadata standard is presented in the following table[27]:

Table 8: Dublin Core Metadata standard vocabulary

No	Element	Element definition
1	Title	A name given to the resource.
2	Creator	An entity primarily responsible for making the content of the resource.
3	Subject	The topic of the content of the resource.
4	Description	An account of the content of the resource.
5	Publisher	An entity responsible for making the resource available.
6	Contributor	An entity responsible for making contributions to the content of the resource.
7	Date	A date associated with an event in the life cycle of the resource
8	Type	The nature or genre of the content of the resource.
9	Format	The physical or digital manifestation of the resource.

[26] L. Steele & T. Orrell (2017). The frontiers of data interoperability for sustainable development. Publish What You Fund and Development Initiatives

[27] Sugimoto, S., Baker, T., & Weibel, S. L. (2002). Dublin Core: Process and Principles. Lecture Notes in Computer Science Digital Libraries: People, Knowledge, and Technology, 25-35.

No	Element	Element definition
10	Identifier	An unambiguous reference to the resource within a given context.
11	Source	A reference to a resource from which the present resource is derived.
12	Language	A language of the intellectual content of the resource.
13	Relation	A reference to a related resource.
14	Coverage	The extent or scope of the content of the resource.
15	Rights	Information about rights held in and over the resource.

Along similar lines, **the interoperability of openly available data will be facilitated through Zenodo**, which adopts community- endorsed practices, since its metadata are stored internally in JSON format according to a defined JSON schema. This encloses HTML microdata that allows machine-readable data to be embedded in HTML documents in the form of nested groups of name-value pairs. Moreover, the JSON schema provides a collection of shared vocabularies in microdata format that can be used to mark-up pages in ways that can be understood by major search engines. In a similar manner, GitHub can facilitate storing metadata either through git notes[28] or by associated description to a branch. The metadata stored can follow the same JSON format. Furthermore, interoperability is key to AIoD platform, which facilitates it through the application of the AIoD Metadata Model, which specifies the entities participating in the European AI ecosystem, the relations between them, and the properties that adequately describe all these entities to the platform's Metadata Catalogue. This ensures that the collected information from participants in the European AI Ecosystem is easily ingested, maintained, updated, and interlinked[29].

COLOSSUS' data will offer qualified references to other data. A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. Our goal is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data, balanced against the time/energy involved in making a good data model. To be more concrete, our references will specify if one dataset builds on another data set, if additional datasets are needed to complete the data, or if complementary information is stored in a different dataset. The links between the datasets will also be described and, all datasets will be properly cited, including their persistent identifiers.

3.4. Increase data re-use

3.4.1 Documentation for validating data analysis and facilitating data re-use

By utilising Zenodo for sharing the project's openly available data, COLOSSUS ensures the facilitation of data access, validation and re-use, in compliance to the general policies of Zenodo regarding content, access and reuse. More specifically, the following principles are followed by Zenodo to make data re-useable according to the FAIR principles[30]:

- R1: (meta)data are richly described with a plurality of accurate and relevant attributes.

Each record contains a minimum of DataCite's mandatory terms, with optionally additional DataCite recommended terms and Zenodo's enrichments.

- R1.1: (meta)data are released with a clear and accessible data usage license.

[28] Git-SCM. Retrieved from: https://git-scm.com/docs/git-notes
[29] AI-on-Demand Platform. "Interoperability is Key: How AIoD facilitates connectivity and ease of information exchange". Retrieved from: https://www.ai4europe.eu/ethics/articles/interoperability-key-how-aiod-facilitates-connectivity-and-ease-information
[30] Retrieved from: https://about.zenodo.org/principles/

License is one of the mandatory terms in Zenodo's metadata and is referring to an Open Definition license. Data downloaded by the users is subject to the license specified in the metadata by the uploader.

- R1.2: (meta)data are associated with detailed provenance.

All data and metadata uploaded is traceable to a registered Zenodo user. Metadata can optionally describe the original authors of the published work.

- R1.3: (meta)data meet domain-relevant community standards.

Zenodo is not a domain-specific repository, yet through compliance with DataCite's Metadata Schema, metadata meets one of the broadest cross-domain standards available.

On the other hand, by utilising GitHub for sharing the project's open-source solutions developed along the datasets used for the various experiments COLOSSUS offers access, and thus increased re-use, to the scientific community that use it as their main knowledge sharing online repository.

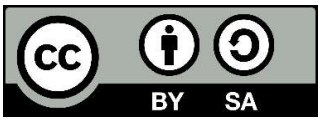
Similarly, AIoD acts as a repository and access point, rather than a traditional data storage archive, namely, while the primary digital content (such as large datasets or model weights) remains on linked host services (e.g., institutional repositories, domain-specific archives, or other data hubs), the metadata catalogue ensures that persistent, structured descriptions and links to these resources are openly available[31]. Thus, also facilitating re-use by the wider community through clearly signposted pathways to access AI-relevant data.

3.4.2 License schemes to permit the widest use possible

Open data will be made freely available in the public domain to permit the widest re-use possible. Moreover, the application of a licence to COLOSSUS open data is a simple way to ensure that any interested third-party can re-use it. In this context, licences are the instrument which permit a third-party to copy, distribute, display and/or modify the project's data only for the purposes that are set by the licence. Licences typically grant permissions on condition that certain terms are met. While the precise details vary, three conditions are commonly found in licences which are the attribution, non-derivative, and non-commerciality.

Along these lines, COLOSSUS publishes openly available data under the **Creative Commons licencing scheme** to foster their re-use and build an equitable and accessible environment for them. Zenodo provides COLOSSUS the **opportunity to publish its open data under five Creative Common licences** as follows:

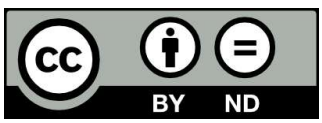
- **Creative commons Attribution-Share Alike 4.0** (CC BY-SA 4.0) according to which any third party can freely copy, distribute, display and modify the datasets for any purpose. Remix, transform, or built upon data, must be distributed under the same license as the original. Third parties must give appropriate credit, provide a link to the license, and indicate if changes were made.



- **Creative Commons Attribution 4.0 International** (CC BY 4.0) according to which any third party can freely copy, distribute, display and modify the datasets for any purpose. Third parties must give appropriate credit, provide a link to the license, and indicate if changes were made.

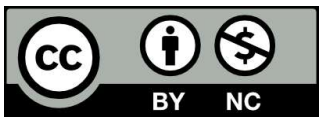


- **Creative Commons Attribution-No Derivatives 4.0 International** (CC BY-ND 4.0) during which any third party can freely copy, distribute, display and modify the datasets for any purpose. Remix, transform, or built upon data, however, must not be distributed. Third parties must give appropriate credit, provide a link to the license, and indicate if changes were made.



[31] AI-on-Demand Platfom LinkedIn. "The Metadata Catalogue: the backbone of AI-on-Demand — Part One" Retrieved from: https://www.linkedin.com/pulse/metadata-catalogue-backbone-ai-on-demand-part-one-aiondemand-sid9f/

- **Creative Commons Attribution-Non-Commercial 4.0 International** (CC BY-NC 4.0) based on which third parties can copy, distribute, display and modify the datasets for any purpose other than commercial unless they get a permission by project partners first. Third parties must give appropriate credit, provide a link to the license, and indicate if changes were made.



- **Creative Commons Attribution-Non-Commercial-No-Derivatives 4.0 International** (CC BY-NC-ND 4.0) according to which third parties can copy, distribute, display and modify the datasets for any purpose other than commercial unless they get a permission by project partners first. Remix, transform, or built upon data, however, must not be distributed. Third parties must give appropriate credit, provide a link to the license, and indicate if changes were made.



Beyond Zenodo licence schemes, Github offers a wider superset of licenses[32] that extends to other types of licenses besides Creative Commons, including but not limited to Academic Free License v3.0, Berkeley Software Distribution (BSD) 2 and 3 licences, European Union Public License 1.1, GNU General Public License v2.0/v3.0, Open Software License 3.0, Educational Community License v2.0, and more.

Moreover, AIoD offers several licensing models that can be used to share user content. These include GPL, APACHE, LGPL, MIT, COMMERCIAL, and Creative Commons. Metadata for uploaded resources is made open under CC0 or equivalent, in alignment with FAIR principles[33].

Different licensing schemes may be selected to better fit the need of COLOSSUS' open data ensuring not only their long-term preservation and re-use but also the interests of the consortium along with the rights of individuals for whom the data are about. In such a case, this subsection of the DMP will be updated accordingly.

3.4.3 Availability for re-use

The re-use of data is a key component of COLOSSUS' methodology for making data FAIR. In fact, making data available for re-use ensures interested stakeholders, other than project partners, can benefit from these data, contributing towards maximising the impact of the project. **Rich metadata** created based on metadata standards that enable proper discovery as well as **appropriate licensing schemes facilitate the re-use of COLOSSUS' open data**, allowing them to find valuable utility even after the end of COLOSSUS project.

In principle, it is expected that data will become available for re-use no later than 120 days after the end of its processing in the framework of the project (i.e., collection, anonymisation, aggregation, etc.) to ensure that any additional data management activities required to this end do not compete with the timely delivery of the project's planned outputs.

With that in mind, the expected time that COLOSSUS' data will be made openly accessible and uploaded to Zenodo is indicatively provided in the following table:

Table 9: Expected time that data will be made open through Zenodo[34]

No	Data	Expected time for making data open	Notes
1	Market analysis	31/05/2026	After completion of D1.1

[32] GitHub. "Managing your repository's settings and features. Retrieved from: https://docs.github.com/en/repositories/managing-your-repositorys-settings-and-features/customizing-your-repository/licensing-a-repository
[33] AI-on-Demand Platform. "Terms and Conditions". Retrieved from: https://aiod.eu/terms-conditions/
[34] This timetable is based on expectations and may be modified during the course of the project taking into account any unforeseen risk that may occur.

No	Data	Expected time for making data open	Notes
2	Identification of regulatory, ethics and security standards requirements	31/05/2026	After completion of D1.1
3	Co-design of the COLOSSUS requirements and system architecture	31/07/2026	After completion of D1.2
4	Pilot Use Case Scenarios Planning	-	Closed data
5	Replication guidelines	31/12/2027	After the approval of D5.3
6	Policy recommendations	31/12/2027	After the approval of D5.3
7	Business development	-	Closed data
8	Collaboration and synergies with relevant projects and initiatives	-	Data will become available as soon as possible after a synergy
9	Setup and Operation of the SAB	28/02/2026	Along the release of the COLOSSUS website.
10	USV data collection, processing and decision making	-	After completion of MS4
11	UAVs and UGVs data collection, processing and decision making	-	Closed data
12	Secure interfaces for port security information exchange	-	Closed data
13	Intelligent Swarm navigation for diverse UV types	-	Closed data
14	3D reconstruction and modelling for underwater inspection	-	Closed data
15	COLOSSUS Web Platform	31/12/2027	After completion of D4.2
16	Pilot Preparation and Deployment	31/12/2027	Stored in AASA secure servers and project repository. Sensitive logs deleted after validation.
17	Technical evaluation and market validation	31/12/2027	Technical KPIs in project repository (open access). Raw sensor data stored internally at AASA.
18	Website analytics	31/12/2027	Public information will be available as needed via COLOSSUS online presence

No	Data	Expected time for making data open	Notes
19	Social media account statistics	31/12/2027	Public information will be available as needed via COLOSSUS online presence
20	Data collected from project events	31/12/2027	Public information will be available as needed via COLOSSUS online presence
21	Newsletter subscription data	-	Closed data
22	Data collected from dissemination and communication activities	31/12/2027	Public information will be available as needed via COLOSSUS online presence

3.4.4 Data provenance

Data provenance is the documentation of where a piece of data comes from and the processes and methodology by which it was produced. Put simply, provenance answers the questions of why and how the data was produced, as well as where, when and by whom[35]. Accurately recording data provenance is a cornerstone of good data management. COLOSSUS will use specific elements of the **Dublin Core Metadata Standards**[36] and the **W3C Provenance Data Model**[37], to generate specific text files (e.g., README) that will accurately capture the history of each data entity throughout its versions (e.g., based on the DOI versioning Zenodo provides)[38].

3.4.5 Data quality assurance processes

Quality Assurance (QA) and **Quality Control** (QC) activities are an integral part of COLOSSUS' data management methodology and are implemented prior to the publication of any data to Zenodo, GitHub or AIoD, safeguarding the transparency, consistency, comparability, completeness and accuracy of the data.

QA is a planned system of review procedures conducted outside the framework of developing a dataset, by personnel not directly involved in the dataset development process[39]. In the context of COLOSSUS, it takes the form of **peer-reviews of methods and/or data summaries** to assess the quality of the dataset and identify any need for improvement, ensuring that the dataset correctly incorporates the scientific knowledge and data generated.

QC is defined as a system of checks to assess and maintain the quality of the dataset being compiled[40]. The relevant procedures of COLOSSUS are designed to provide routine technical checks as they measure and control data consistency, integrity, correctness and completeness as well as identify and address errors and omissions. In this context, QC checks cover everything from data acquisition and handling, application of approved procedures and methods, and documentation. Some of the general quality checks undertaken in the framework of the project include checking (i) for transcription errors in data input; (ii) that scale measures are within the range of acceptable values; and (iii) whether proper naming conventions are used.

[35] https://ardc.edu.au/resource/data-provenance/
[36] https://www.dublincore.org/resources/userguide/creating_metadata/#Provenance
[37] https://www.w3.org/TR/prov-dm/
[38] https://help.zenodo.org/
[39] 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Vol. 1 General Guidance and Reporting, CHAPTER 6 Quality Assurance / Quality Control and Verification.
[40] 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Vol. 1 General Guidance and Reporting, CHAPTER 6 Quality Assurance / Quality Control and Verification.

4. Other research outputs

The implementation of COLOSSUS entails activities that will generate or re-use other research outputs. This includes digital (e.g., software, workflows, protocols, models) or physical (e.g., Unmanned Oceanic Patrol Vessel (UOPV)) research outputs. The management of these other research outputs in the context of COLOSSUS will be in line with FAIR data principles, as concisely outlined in the table that follows.

Table 10: Other research outputs

Task	WTL	Responsible Partner	Research output	Brief description	Digital/ Physical	Reasons	Benefits/exploitation	Interested stakeholders	Format	Size (KB)	Repository	Open access	FAIR principles	Timing for making data available
T2.1	TV	TV	UOPV	Upgraded and integrated Unmanned Oceanic Patrol Vessel (UOPV) platform, incorporating secure communication modules, AI-enabled perception systems, and swarm coordination capabilities for port security and resilience operations. The UOPV constitutes background IP owned by Tecnoveritas and is technically enhanced within COLOSSUS.	Physical (with embedded digital components)	Core operational platform required to validate autonomous perimeter surveillance, swarm-enabled patrol intelligence, and system integration under real port conditions.	Commercial deployment in EU ports and critical maritime infrastructures; integration into Tecnoveritas' product portfolio; replication in security and maintenance applications; contribution to COLOSSUS business plan and go-to-market strategy.	Public and Private actors active in port management and security	PDF when appicable	N/A	Technical documentation stored in secure consortium repository (restricted access – Google Drive / internal secured storage).	NO	YES	N/A

T2.3	iNGS	iNGS	Secure API	The Secure API supports interoperabity with external systems by enabling secure, reliable, and trustworthy data exchange, with access via security tools to the port community.	Digital	The Secure API enables interoperability with external systems by providing standardized, encrypted endpoints through robust authentication protocols like OAuth 2.0 and TLS 1.3 encryption, preventing unauthorized access and data breaches during exchanges.		Public and Private actors active in port management and security	Deliverables, Executable SW	>500	Project Git	NO	N/A	N/A
T3.2	iNGS	iNGS	Intelligent Swarm navigation Algorithm	Intelligent cooperative control and navigation of diverse types of UVs to collectively detect, and map the area under consideration, considering the various capabilities in terms of sensing, scope, and movement of the different subsystems.	Digital				Executable SW	>100	Project Git	NO	YES	N/A

5. Allocation of resources

5.1. Estimated costs for making data FAIR

The costs required for making the data collected/generated during COLOSSUS' activities FAIR are integrated into the budget of the project. With that in mind, the table which follows provides an overview of the estimated costs of making data FAIR as well as their budget source within the framework of COLOSSUS.

Table 11: Estimated costs for making data FAIR

#	Data Processing / Management Activity	Budget source	Total estimated effort in Person Months[41]	Total estimated cost in Euro[42]
1	Collection	Budget allocated to the WP under which the respective data are processed	19,44	86.855,86 €
2	Documentation	Budget allocated to the WP under which the respective data are processed	5,55	24.815,96 €
3	Storage	Budget allocated to the WP under which the respective data are processed	2,78	12.407,98 €
4	Access and security	Budget allocated to the WP under which the respective data are processed	2,78	12.407,98 €
5	Preservation	Budget allocated to the WP under which the respective data are processed	1,39	6.203,99 €
6	Availability and re-use	Budget allocated to the WP under which the respective data are processed	8,33	37.223,94 €
7	Overall data management	WP8,9	0,96	4.289,40 €
		Total	**41,23**	**184.205,11 €**

In order to produce the estimations of the costs for making data FAIR in the context of COLOSSUS, a series of **assumptions** were made, taking into account the respective **guidelines** provided by the Research Data

[41] The total estimated effort for each data processing/management activity reflects the cumulative effort for the implementation of the respective activity for all data collected/generated across the different WPs of COLOSSUS.
[42] The total cost of each data processing/management activity is calculated by multiplying the effort estimated for the respective activity with the weighted average cost of a person month in the framework of COLOSSUS.

Management Support, a multidisciplinary network of data experts within Utrecht University[43], as well as of the UK Data Service and its data management costing tool[44]. With that in mind, the estimated costs for making COLOSSUS' data FAIR cover **data-related activities and resources across the data lifecycle**, spanning from collection and documentation through storage and preservation over to sharing and re-use.

In particular, costs for **data collection** cover activities necessary for acquiring external datasets (if required), gathering/generating new data, transcribing (if applicable), formatting and organising these data as well as acquiring informed consent from data subjects. This data processing activity reflects the majority of the costs required for making data FAIR as the majority of COLOSSUS' data constitutes new data collected/generated over the course of the project. At the same time, **data documentation** costs address the effort required for describing data (e.g., marking data with variable and value labels, code descriptions, etc.) as well as creating well-defined metadata along with a meaningful description of the context and methodology of how data was collected/generated and processed (where necessary).

Costs for **data storage** include the resources required for ensuring adequate storage space for the data as well as the effort necessary for conducting data back-ups, while **data access and security** costs encompass costs related to ensuring access to the data as well as for protecting it from unauthorised access or use or from disclosure. Given that the storage of COLOSSUS' data will not require the procurement of additional space (other than what is already available to project partners) as well as that no special measures or software are required to access and secure the data (other than what is inherently built into the repositories of COLOSSUS' data), such costs are kept to a minimum.

Data preservation costs, on the other hand, are estimated relatively higher than data storage, access and security costs, as additional effort will be required in several cases in order to convert the collected/generated data from their original form (e.g., physical interview transcripts) to an open and/or machine-readable format suitable for long-term preservation (e.g., to an .xlsx format.). Adequate effort for **data availability and re-use** costs is also foreseen to safeguard the appropriate digitisation and anonymisation of the data as well as cover any resources required for data sharing and cleaning. Along the same lines, appropriate effort is foreseen for **overall data management** as well, in order to cover the effort related with the operationalisation of data management in the framework of COLOSSUS.

Finally, costs for **long-term preservation** in the framework of COLOSSUS are assumed to be negligible since the open data of the project will be hosted in the repository of Zenodo, GitHub or AIoD free of charge.

5.2. Data management responsibilities

For the effective, proper and secure handling of the data collected/generated in the frame of COLOSSUS, specific data management roles have been established within the data management methodology and procedures of the project. These responsibilities are outlined in this section of the DMP and are as follows.

Project Coordinator (PC): responsible for overall data management in the framework of COLOSSUS. At the same time, the PC is responsible for the templates for the Informed Consent Form and the Data Subject Request Form (Annex II and III) which will be appropriately adjusted and utilised by project partners (under their responsibility) during the relevant activities of the project as well as for the project's Privacy Policy (Annex I) that has been uploaded on the project's website. The PC in collaboration with the relevant project partners (e.g., Task Leaders) will examine if additional specific privacy policies are required for certain project's tasks and will coordinate the elaboration of such privacy policies. Finally, the PC coordinates with Work Package Leaders and Co-leaders, Task Leaders, and Responsible Partners to determine whether and how the data collected /

[43] Research Data Management Support. Guides: Costs of data management. Utrecht University. Retrieved from: https://www.uu.nl/en/research/research-data-management/guides/costs-of-data-management

[44] UK Data Service. Costing Data Management. Retrieved from: https://www.ukdataservice.ac.uk/manage-data/plan/costing

generated by the project are shared and become available for re-use, contributes to its quality assurance and uploads the project's openly available data to Zenodo, GitHub, and AIoD.

Work Package Leaders (WPL): The WPL is responsible for coordinating the implementation of the data processing activities performed under the WPs they are leading. Moreover, they align with the respective Work Task Leader on whether and how the data gathered/produced under the tasks that fall within the WP they are leading will be shared and/or re-used. This includes the definition of access procedures as well as potential embargo periods along with any necessary software and/or other tools which may be required for data sharing and re-use. Finally, the WPL are the main responsible for assuring the quality of the data stemming from the activities of the WP they are leading, including assessing their quality and indicating any need for improvement to the respective Work Task Leaders.

Work Task Leaders (WTL): WTLs are responsible for the data collected/generated or re-used in the frame of the tasks that fall under their leadership as well as for safeguarding their appropriate and timely processing. Moreover, they are responsible for properly adjusting the Informed Consent Form and Data Subject Request Form templates, to the needs and specificities of the activities carried out in the task they are leading. WTLs are responsible for identifying the need for a specific privacy policy regarding the task they are leading and collaborate with the PC for drafting and releasing it to the public. Finally, they undertake any necessary actions to prepare the data collected/generated or re-used through the tasks they are leading for sharing either within the consortium or openly (including the use of proper naming conventions, application of suitable anonymisation techniques, creation of appropriate metadata and documentation, etc.).

Partners: All project partners are tasked to collect, digitise, anonymise, store, destroy and/or otherwise process data for the specific purpose of the activity in which it has been collected/generated or re-used within the project. They are responsible for appropriately collecting the necessary consent for processing data as well as for ensuring that the Informed Consent Form and the Data Subject Request Form used to this end are properly adjusted to the needs of the activity they are participating (including references to the project's Privacy Policy and any other applicable specific privacy policies) and, in any particularities, applicable to their organisation while ensuring adherence to provisions of relevant national data protection legislation in their respective country. Moreover, they are responsible for managing the consents they have collected with a view to demonstrating their compliance with the relevant applicable EU and national regulation(s). Finally, they perform quality checks to assess and maintain the quality of the dataset(s) held within their records.

Data repositories: Data repositories are tasked with the storage and long-term preservation of the project's data. In this respect, Zenodo, GitHub, and AIoD will maintain and preserve the openly available data of COLOSSUS, enabling its sharing and re-use. To this end, Zenodo assigns metadata and DOIs to the data, while also taking all necessary measures to securely back-up the data and restore it, safeguarding its long-term preservation. Accordingly, metadata will be included in the GitHub publicly available data, with state-of-the-art security ensured[45], while AIoD's Metadata Catalogue unifies AI research metadata across platforms such as Zenodo and GitHub, with data stored under a unified Metadata Schema. This means that the metadata about a dataset on the platform will always have the same structure, regardless of where it was originally hosted.

In this context, the following table illustrates the allocation of data management responsibilities amongst the members of the COLOSSUS consortium per data collected/generated or re-used under each WP.

[45] GitHub. "Security". Retrieved from: https://github.com/security

Table 12: Data management responsibilities per data collected/generated under each WP

WP	WPL	Data	Tasks	WTL	Responsible Partners
WP1	UNINOVA	Market analysis	Task 1.1	Q-PLAN	ALL
		Identification of regulatory, ethics and security standards requirements	Task 1.2	iNGS	ALL
		Co-design of the COLOSSUS requirements and system architecture	Task 1.3	UNINOVA	ALL
		Pilot Use Case Scenarios Planning	Task 1.4	AASA	ALL
WP2	TV	USV data collection, processing and decision making	Task 2.1	TV	UNINOVA iNGS AASA
		UAVs and UGVs data collection, processing and decision making	Task 2.2	UNINOVA	iNGS AASA
		Secure interfaces for port security information exchange	Task 2.3	iNGS	UNINOVA AASA
WP3	iNGS	Multi-modal data-efficient harmonisation	Task 3.1	UNINOVA	iNGS, TV
		Intelligent Swarm navigation for diverse UV types	Task 3.2	iNGS	UNINOVA, TV
		3D reconstruction and modelling for underwater inspection	Task 3.3	UNINOVA	TV, iNGS
WP4	TV	Integration lab testing and fine tuning	Task 4.1	TV	UNINOVA iNGS AASA
		Lab stress-testing and technology validation	Task 4.2	TV	UNINOVA iNGS AASA
		COLOSSUS Web Platform	Task 4.3	TV	ALL
WP5	AASA	Pilot Preparation and Deployment	T5.1	AASA	TV, UNINOVA, iNGS
		Technical evaluation and market validation	T5.2	AASA	Q-PLAN, TV

WP	WPL	Data	Tasks	WTL	Responsible Partners
		Replication guidelines	**T5.3**	AASA	UNINOVA, Q-PLAN
		Policy recommendations	T5.4	AASA	UNINOVA, Q-PLAN
WP6,7	Q-PLAN	Social media account statistics	Task 6.1/7.1	Q-PLAN	Q-PLAN
		Data collected from project events	Task 6.1/7.1	Q-PLAN	ALL
		Newsletter subscription data	Task 6.1/7.1	Q-PLAN	Q-PLAN
		Data collected from dissemination and communication activities	Task 6.1/7.1	Q-PLAN	ALL
		Business development	Task 6.3/7.3	Q-PLAN	Q-PLAN
		Collaboration and Synergies with relevant projects and Initiatives	Task 6.5/7.5	Q-PLAN	ALL
WP8,7	TV	Setup and Operation of the Security Advisory Board	Task 8.2/9.2	UNINOVA	ALL

6. Data Security

COLOSSUS will securely handle any collected/generated or re-used data throughout its entire lifecycle as it is essential to safeguard these data against accidental loss and/or unauthorised access. To achieve this the project will apply appropriate technical and organisational measures based on a risk assessment of the relevant data that takes into account the impact and the likelihood of a potential data breach. With that in mind, the project's data security strategy aims at minimizing the probability that a data breach will occur during the course and after the completion of COLOSSUS, resulting either from human error or hardware failure, as well as inhibit any unauthorised access. Particularly, in case of personal data collection/generation it is crucial that these **data can only be accessible by those authorised to do so.**

Regarding the security and handling of **sensitive data** deriving **from the use cases' demonstration**, clear definition of data location at all stages. Before operations, mission plans and protocols are stored on AASA's secure internal servers. After deployment, raw data is captured locally and moved to restricted shipyard storage for validation. Only sanitized, non-sensitive results are then transferred to the project repository. This 'Before-and-After' storage clarification ensures full control over sensitive infrastructure data throughout its lifecycle.

All project partners are responsible for processing[46] data using appropriate means, such as private servers or cloud service providers that adhere to the relevant legal data protection requirements (e.g., GDPR) and will ensure that these **data are protected,** and any **necessary data security controls have been implemented**, to minimize the risk of information leak and destruction. This case refers to the data that will be closed and therefore will not be shared and/or re-used within the framework of the project. In this case, to minimize the consequences of potential data losses, the data will be **backed up at regular time intervals based on change frequency and criticality. The backed-up files will be stored in appropriate storage media including external hard drives, flash drives, Network-Attached Storage (NAS) devices and reputable cloud services**, so as to safeguard their preservation, while also enabling their recovery at any time. Moreover, **integrity checks**[47] will be carried out regularly ensuring that the stored data has not been changed or corrupted.

Access to closed data will only be permitted to authorised project partners. In case there is a **personal data breach**, the responsible **project partner will notify, without undue delay** and, where feasible, no later than **72 hours after having become aware of it, its competent national supervisory authority** (e.g., data protection authority) **as well as the data subject(s) that may be affected by the breach.** Moreover, the responsible partner will document any personal data breaches, including information such as the facts relevant to the breach, its effects and the remedial action(s) taken.

Identification and authentication access controls play an important role in the context of the project, as they help partners to protect the data collected/generated or re-used during COLOSSUS and especially personal data. To this end, each project partner is responsible for and committed to ensuring the application of appropriate access controls to the data they are processing. Finally, in order to safeguard the privacy of the users of the COLOSSUS website, a dedicated **privacy policy** will define the way in which this online space collects, processes and uses personal data, the security procedures followed, the users' rights as well as the cookies policy employed.

On another note, openly available data will be stored safely for long-term preservation on Zenodo, in the same cloud infrastructure as research data from CERN's Large Hadron Collider, using CERN's battle-tested repository software INVENIO, which is used by some of the world's largest repositories (such as INSPIRE HEP and the

[46] Processing, according to Regulation (EU) 2016/679 of the European Parliament (General Data Protection Regulation), means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.
[47] An integrity check is the process of comparing the current state of stored data and/or programs to a previously recorded state in order to detect any changes.

CERN Document Server). Along these lines, data are stored and backed-up in CERN's EOS service in an 18 petabytes disk cluster. Both data files and metadata are kept in multiple online replicas and independent replicas ensuring their long-term preservation as well as their recovery when necessary. Moreover, for each file two independent MD5 checksums are stored. One checksum is stored by INVENIO, used to detect changes to files made from outside of it whereas the other checksum is stored by EOS, and used for automatic detection and recovery of file corruption on disks. In this context, access control is applied by the different level of openness that Zenodo allows (i.e., open, restricted and closed).

Accordingly, data stored at GitHub can also be considered to be stored safely for long-term preservation. In general, GitHub is a GDPR compliant platform that ensures data privacy and security. GitHub is a Trusted Cloud Provider (TM) with the Cloud Security Alliance (CSA)[48], whereas its Information Security Management System (ISMS) has been certified against the ISO/IEC 27001:2013 standard.

Similarly, AIoD is a fully GDPR compliant platform that guarantees users their right in line with the General Data Protection Regulation 2016/679. To complement this, it implemented several measures that ensure data security, such as strong password policies in all servers, HTTPS protocol for interacting with APIs and Web Clients, Secure Shell (SSH) protocol for server connection, and periodical server updates with latest security fixes[49]. AIoD ensures the highest standards for data protection, compliance, and operational resilience by utilising a sovereign cloud infrastructure — with secure DevOps and full data sovereignty which constitutes a secure and independent infrastructure in line with all EU requirements[50].

[48] Cloud Security Alliance. Retrieved from: https://cloudsecurityalliance.org/
[49] AI-on-Demand Platform. "Privacy Policy". Retrieved from: https://aiod.eu/privacy-policy/
[50] Convotis. "European Commission: AI Platform for Europe – Sovereign, Scalable, Future-Proof". Retrieved from: https://www.convotis.com/en/news/customer-success-eu-commission-aiod-platform/

7. Ethics and other issues

This Chapter addresses the ethical aspects of the COLOSSUS' Data Management Plan and the ethical compliance of the underlying data foreseen to be collected/generated or re-used under the project's activities. The project will process **data that is not included in any special category of personal data** (i.e., non-sensitive data), according to the relevant data protection legislation (e.g., GDPR), **whereas any personal data** related to the subjects of the use cases will be handled according to GDPR requirements and all, potential additional National regulations, by the pilot responsible partners.

Non-sensitive data:

In accordance with the **Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (GDPR),** all personal data processed for project's activities shall be:

- processed lawfully, fairly and in a transparent manner in relation to the data subject;
- collected for specified, explicit and legitimate purposes relative to project's objectives 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 personal data are processed;
- processed in a manner that ensures appropriate security of the personal data (see section 6).

For all personal data processing activities within the framework of the project at least one lawful basis as of Art. 6 GDPR applies. Where informed consent is chosen as the lawful basis for processing, all relevant provisions of the data protection legislation (e.g., Art.7 GDPR) shall be fulfilled. Under this light, further details about the **scope of the activities that entail data collection/generation or re-use** in the frame of COLOSSUS along with the procedures for identifying/recruiting suitable stakeholders to take part in them as well as for obtaining their informed consent are defined by the respective WP Leaders (UNINOVA for WP1, TV for WP2, iNGS for WP3, TV for WP4, AASA for WP5, Q-PLAN for WP6&WP7, TV for WP8&9). Moreover, **personal data processing carried out by partners are in line with relevant EU and national regulations**. The project's Privacy Policy and the templates of the Informed Consent Form and the Data Subject Request Form, used in the implementation of the project's activities, are compliant with the General Data Protection Regulation and annexed to this DMP (see Annex I, II and III). Last but not least, **no transfer of personal data outside the EU is foreseen as part of the project's implementation.** In case of data storage providers situated both inside and outside the EEA, partners are committed to ensure their compliance with the relevant GDPR requirements before they start using the data storage providers' services.

It is important to highlight that **each partner is responsible for ensuring that the templates for the Informed Consent Form and Subject Data Request Form (including references to the project's Privacy Policy and any other applicable specific privacy policies) are appropriately adjusted** according to (i) the needs of the activity for which they are being used by them as well as to (ii) the relevant data protection laws and regulations applicable on European, national and organization level. **All partners should keep records to demonstrate that data subjects have consented to the processing of their personal data** and ensure that consent management mechanisms are in place according to Art. 7(3) of GDPR to make it easy for individuals to withdraw their consent.

Sensitive data:

In general, the consortium will carry out the action in compliance with:

- ethical principles (including the highest standards of research integrity) and
- applicable EU, international and national laws and regulations, including, but not limited to the EU Charter of Fundamental Rights, the European Convention for the Protection of Human Rights and Fundamental Freedoms[51] and its Supplementary Protocols, and the International Council for Harmonisation (ICH) principles[52] for quality, safety, efficacy and good clinical practice where applicable.

The consortium will also pay particular attention to the principle of proportionality, the right to privacy, the right to the protection of personal data, the right to the physical and mental integrity of persons, the right to non-discrimination, the need to ensure protection of the environment and high levels of human health protection. The beneficiaries will ensure that the activities under the action have an exclusive focus on civil applications. And that the activities under the action do not:

- aim at human cloning for reproductive purposes;
- intend to modify the genetic heritage of human beings which could make such modifications heritable (with the exception of research relating to cancer treatment of the gonads, which may be financed);
- intend to create human embryos solely for the purpose of research or for the purpose of stem cell procurement, including by means of somatic cell nuclear transfer, or
- lead to the destruction of human embryos (for example, for obtaining stem cells).

In addition, the beneficiaries will respect the **fundamental principle of research integrity** — as set out in the European Code of Conduct for Research Integrity[53]. This implies compliance with the following principles:

- reliability in ensuring the quality of research reflected in the design, the methodology, the analysis and the use of resources;
- honesty in developing, undertaking, reviewing, reporting and communicating research in a transparent, fair and unbiased way;
- respect for colleagues, research participants, society, ecosystems, cultural heritage and the environment;
- accountability for the research from idea to publication, for its management and organisation, for training, supervision and mentoring, and for its wider impacts and means that beneficiaries must ensure that persons carrying out research tasks follow the good research practices including ensuring, where possible, openness, reproducibility and traceability and refrain from the research integrity violations described in the Code.

Activities raising ethical issues will comply with the additional requirements formulated by the ethics panels (including after checks, reviews or audits as described in Article 25 of the DoA). Before starting an action task raising ethical issues, the beneficiaries will have obtained all approvals or other mandatory documents needed for implementing the task, notably from any (national or local) ethics committee or other bodies such as data protection authorities. The documents will be kept on file according to European and national requirements for data storage and will be submitted upon request by the coordinator to the granting authority.

[51] https://www.echr.coe.int/documents/d/echr/convention_eng
[52] ICH principles:
- https://www.ich.org/page/ich-guidelines
- https://www.ema.europa.eu/en/ich-e6-r2-good-clinical-practice-scientific-guideline

[53] https://allea.org/wp-content/uploads/2023/06/European-Code-of-Conduct-Revised-Edition-2023.pdf/

8. Conclusions

This initial version of the COLOSSUS DMP aims at safeguarding the sound management of the data collected, processed and/or generated during the project's activities across their entire lifecycle, while also making them FAIR. It describes all the underlying processes of the COLOSSUS data management, collection, process and generation, in accordance with the GDPR guidelines, and sheds light on (i) the data being collected, processed, generated and/or re-used under the project activities, (ii) the specific objectives under which each dataset is collected, processed, generated and/or re-used, (iii) the management of the other research outputs of the project (iv) the allocation of resources and data management responsibilities and (v) the data security and ethical aspects of the data.

In the framework of COLOSSUS, the DMP is a living document and is updated throughout the course of the project, considering its latest developments and available results. It is expected to be further developed and updated at least twice by the end of the project (i.e., as D8.3 by M12 and as D9.1 by M24). If necessary, additional ad hoc updates may be released in order to include new data, better detail and/or reflect modifications in the methodologies applied or other aspects relevant to data management (such as costs for making data FAIR, size of data, etc.), changes in consortium policies and plans or other potential external factors.

Annexes

Annex I – Privacy Policy

Within this Annex, the reader can find the project's overall Privacy Policy, which is uploaded on the project's website[54]. Task leaders are responsible for developing any additional privacy policy needed in their tasks and activities in close collaboration with the PC. Moreover, they are responsible for appropriately adjusting and translating the templates to fit the needs and specificities of their task's activities.

1. Who we are

COLOSSUS, a Horizon Europe Innovation Action project, is funded by the European Union, under Grant Agreement No. 101120706, with a budget of approximately €1.5 million. The project started in November 2025 and will last 24 months.

COLOSSUS presents an integrated solution for digital security, resilience, and maintenance of port infrastructures, providing real-time data access from autonomous vehicles and sensors. Developed in response to the increasing demand for maritime security, COLOSSUS aims to revolutionize the protection of EU maritime infrastructures, aligning with the updated EU Maritime Security Strategy (EUMSS 2023). By bringing together Small and Medium-sized Enterprises (SMEs) from various Member States, the project leverages their innovation potential to tackle key challenges outlined in the EU's Blue Growth strategy and the European Green Deal.

The advanced COLOSSUS solutions will be demonstrated in an operational environment (integrated pilot system level) for iterative performance evaluation and validation. Starting with stress testing in a laboratory setting, the project will move towards real-world operational conditions in ports. Defined use cases will enable operational evaluation and validation across diverse scenarios at the AASA naval base in Portugal.

The project partners of the COLOSSUS consortium, listed below, process certain types of personal data for the purposes of the project. Each partner is responsible for the personal data they collect and process during their activities under the framework of the project:

- **TV** | TECNOVERITAS SERVIÇOS DE ENGENHARIA E SISTEMAS TECNOLÓGICOS LDA (project coordinator), Portugal, https://www.tecnoveritas.net/
- **UNINOVA** | UNINOVA-INSTITUTO DE DESENVOLVIMENTO DE NOVAS TECNOLOGIAS-ASSOCIACAO, Portugal, https://www.uninova.pt/
- **iNGS** | INTELLIGENT NEXT GENERATION SYSTEMS LYSEIS PLIROFORIKIS KAI EPIKOINONION SA, Greece, https://ings.gr/
- **Q-PLAN** | Q-PLAN INTERNATIONAL ADVISORS PC, Greece, https://qplan-intl.gr
- **ABB EUOPC** | ABB SRO, Czechia, https://new.abb.com/cz/en/o-nas/technology-and-innovation/operation-center-europe/introduction-operation-center
- **AASA** | ARSENAL DO ALFEITE SA, Portugal, https://www.arsenal-alfeite.pt/

For further information, we can be contacted at: info@colossus-project.eu.

2. How we collect your personal data

We collect personal data both directly and indirectly:

- We obtain personal data directly from individuals in a variety of ways, including the following cases:
  - an individual subscribes to our newsletter/s;
  - an individual registers to attend meetings, workshops and other events we host and during attendance at such events;
  - an individual participates in an interview or survey organised by us;

[54] https://colossus-project.eu/privacy-cookies-policy/

- we establish cooperative relationships with an individual;
- we provide professional services pursuant to our contract with the European Commission;

- We obtain personal data indirectly about individuals from a variety of sources, including:
  - our research partners;
  - our networks and contacts;
  - public and open data sources such as public registers, news articles and internet searches;
  - social and professional networking sites (e.g., LinkedIn).

3. What types of data we collect

We only collect the data necessary for our project's smooth implementation. These data fall into the following categories:

- contact details (name/ surname, e-mail address, phone number);
- professional information (job title, organisation, field of expertise);
- demographics (e.g. age, gender, nationality);
- information about what a person knows or believes;
- videos and photos (from people that attend our events, webinars, workshops etc.).

4. Bases of lawful processing

We process personal data on the following legal bases:

- Legal obligations – for processing activities required for compliance both with applicable national and European legislation as well as with the specific legal and regulatory framework of the Horizon Europe Framework Programme for Research and Innovation of the European Union.
- Consent – for processing activities such as organisation of meetings, workshops, other events, and dissemination of project's results.
- Contractual obligations – for processing activities such as reporting to the European Commission and complying with project's publicity obligations.

5. What we do with your personal data

We process your personal data with the purpose of:

- Disseminating our project's results to different types of stakeholders;
- Sending invitations and providing access to guests attending our meetings, events, and workshops;
- Administering, maintaining, and ensuring the security of our information systems, applications, and websites;
- Processing online requests or queries, including responding to communications from individuals;
- Complying with contractual, legal, and regulatory obligations.

6. How we secure your personal data when we process it

We continuously apply a personal data risk assessment process to identify, analyse, and evaluate the security risks that may threaten your personal data. Based on the results of this risk assessment, we define and apply a set of both technical and organizational measures to mitigate the above security risks, including but not limited to:

- Data Protection Policies to guide personnel involved in the COLOSSUS project when processing your data;
- Written contracts with organisations that process personal data on our behalf;
- Non-Disclosure Agreements with personnel of the COLOSSUS consortium partners;
- Back up process, antimalware protection, access control mechanisms, etc.
- Some of COLOSSUS partners have appointed a Data Protection Officer.

7. Do we share personal data with third parties?

Your personal data may be shared with trusted third parties to help us deliver efficient and quality services. When we do so, we ensure that recipients are contractually bound to safeguard the data we entrust to them before we share the data. We may engage with several or all of the following categories of recipients:

- Parties that support us as we provide our services (e.g., cloud-based software services such as Dropbox, Microsoft SharePoint, Google);
- Our professional advisers, including lawyers, auditors, and insurers;
- Dissemination service providers (e.g., MailChimp);
- Law enforcement or other government and regulatory agencies or other third parties as required by, and in accordance with applicable law or regulation;
- The European Commission according to our relevant contractual obligations.

8. Do we transfer your personal data outside the European Economic Area?

We do not own file servers located outside the European Economic Area (EEA). However, some partners may use cloud and/or marketing services from reputable providers such as SharePoint, DropBox, MailChimp, Google, etc., situated both inside and outside the EEA. We always check that such providers comply with the relevant GDPR requirements before starting to use their services.

9. Do we use cookies?

Our website, https://colossus-project.eu/, uses cookies. We use cookies to improve site navigation and to analyse our traffic. In the next paragraphs, you will learn what cookies are, what cookies we use and how you can disable their use.

Cookies are small alphanumeric characters files that are transferred from websites that you visit to your PC, smartphone, tablet or any other electronic device that you use for browsing. Cookies assign your computer with a unique ID, which in turn becomes your own identity each time you return to a website. Cookies do not damage your electronic equipment, nor can they read information from other files on your computer's hard drive.

Websites use cookies to "remember" for a period your actions and preferences such as language, font size and other display features. In this way, you are not required to enter these preferences each time you visit a website and navigate its pages. Furthermore, cookies help website owners to analyse how you are using their website and whether or not you are facing problems as you navigate them.

Personal Data Protection legislation states that we can store cookies on your device if they are strictly necessary for the operation of our website. For all other types of cookies, we need your consent. You can change or withdraw your consent at any time from the Cookies Settings on our website.

On our website, https://colossus-project.eu/, we use necessary, statistics and functional cookies. Some cookies are placed by third-party services (e.g., Google Analytics).

Without the necessary cookies, the website cannot perform its basic functions. These cookies are generally placed automatically either when a webpage is loaded or as a result of a user's request (that cannot be fulfilled without using a relevant cookie). Usually, necessary cookies expire when you close your browser.

Statistics cookies help us to measure our website's traffic and where it comes from so that we can improve its performance. Moreover, they collect information about which pages are more popular and how visitors use our website. None of this information can be used to identify you. It is all aggregated and, therefore, anonymized. This category includes cookies from third-party analytics services such as Google Analytics. Our website uses the following statistics cookies:

Name	Provider	Purpose	Expiry
_ga	Google Analytics	This cookie is used to distinguish unique users by assigning a randomly generated number as a client identifier. It is included in each page request on a site and used to calculate visitor, session and campaign data for the site's analytics reports.	2 years

Name	Provider	Purpose	Expiry
_gid	Google Analytics	This cookie stores and updates a unique value for each page visited. It is used to generate statistical data on how the visitor uses the website. It collects anonymized data including the number and origin of visitors and the websites they visited.	1 day
_gat	Google Analytics	This cookie is used to throttle the request rate, that is limiting the collection of data on high-traffic sites.	1 minute

Functionality cookies allow websites to remember the user's site preferences and choices they make on the site including username, region, and language. This allows the website to provide personalized features if you share your location. They are anonymous and do not track browsing activity across other websites.

You can delete cookies and stop them from being installed in your browser. There is no standardized way to remove cookies since different browsers clear cookies using different procedures. Please follow the instructions provided by each browser manufacturer on how to remove cookies.

If you wish to stop sending your information to Google Analytics, you may install in your browser the Google Analytics Opt-out Browser Add-on.

You may select and change the cookie settings (with the exception of the technically necessary cookies) by clicking on Cookie settings. By making a cookie active, you consent to its use according to the provisions of this Cookies Policy.

10. Your rights

You have the following rights regarding our processing of your personal data:

- Right to withdraw consent – You can withdraw consent that you have previously given to one or more specified purposes to process your personal data. This will not affect the lawfulness of any processing carried out before you withdraw your consent.
- Right of access – You can ask us to verify whether we are processing personal data about you and if so, to have access to a copy of such data.
- Right to rectification and erasure – You can ask us to correct our records if you believe they contain incorrect or incomplete information about you or ask us to erase your personal data after you withdraw your consent to processing or when we no longer need it for the purpose it was originally collected.
- Right to restriction of processing – You can ask us to temporarily restrict our processing of your personal data if you contest the accuracy of your personal data, prefer to restrict its use rather than having us erase it, or need us to preserve it for you to establish, exercise or defend a legal claim. A temporary restriction may apply while verifying whether we have overriding legitimate grounds to process it. You can ask us to inform you before we lift that temporary processing restriction.
- Right to data portability – In some circumstances, where you have provided personal data to us, you can ask us to transmit that personal data (in a structured, commonly used, and machine-readable format) directly to another entity.
- Right to object – You can object to our use of your personal data for direct marketing purposes, including profiling or where processing has taken the form of automated decision-making. However, we may need to keep some minimal information (e.g., e-mail address) to comply with your request to cease marketing to you.
- Right to make a complaint to your local Data Protection Authority (DPA) (see https://ec.europa.eu/justice/article-29/structure/data-protection-authorities/index_en.htm) regarding any concerns you may have about our data handling practices.

To ask us to do anything of the above, you can contact us by email: info@colossus-project.eu. We will promptly examine your request against the relevant requirements of the laws and regulations governing privacy and personal data protection and we will answer the latest within 30 days after receiving your request. We will ask you for some kind of identification (e.g. photocopy of your identity card or passport) to avoid the non-authorized reveal of your personal data. If for reasons of the complexity of the request or a multitude of requests, we are

unable to respond promptly, we will notify you within 30 days of any delay, which in no case may exceed two months from the expiration of the 30-day deadline.

11. How we use your information?

- We use the information we collect in various ways, including to:
- Provide, operate, and maintain our website
- Improve, personalise, and expand our website
- Understand and analyse how you use our website
- Develop new products, services, features, and functionality
- Communicate with you, either directly or through one of our partners, including for customer service, to provide you with updates and other information relating to the website, and for marketing and promotional purposes
- Send you emails
- Find and prevent fraud

12. Log Files

Colossus follows a standard procedure of using log files. These files log visitors when they visit websites. All hosting companies do this as a part of hosting services' analytics. The information collected by log files include internet protocol (IP) addresses, browser type, Internet Service Provider (ISP), date and time stamp, referring/exit pages, and possibly the number of clicks. These are not linked to any information that is personally identifiable. The purpose of the information is for analysing trends, administering the site, tracking users' movement on the website, and gathering demographic information.

13. How long do we retain personal data?

We retain personal data to provide our services, stay in contact with you and comply with applicable laws, regulations, and contractual obligations to which we are subject. Please note that we have an obligation to retain data concerning projects funded by the Horizon Europe Framework Programme for Research and Innovation of the European Union for up to five years after the end of the project (unless auditors request further retention). After the expiry of the retention period, and unless further legitimate grounds for retention arise, we will dispose of personal data in a secure manner.

14. Disclaimer of liability for third-party websites

Although our site may contain links to third-party sites, including the sites of the COLOSSUS consortium partners, we are not responsible for the privacy practices or content of these sites, and we expressly disclaim any liability for any loss or damage that may be caused by the use of these links. We do not monitor the privacy practices or the content of these sites. If you have any questions about the privacy practices of another site, you should contact the site's responsible personnel. We suggest you read the privacy policy of each website you interact with, before allowing the collection and use of your personal data.

We may also provide social media features that allow you to share information on your social networks and interact with the COLOSSUS project on various social media sites. The use of these social media features may result in the collection or sharing of information about you. We recommend that you check the privacy policies and regulations of the social networking sites you interact with to ensure that you understand what information may be collected, used and disclosed by these sites.

15. Children

We do not knowingly collect, use, or disclose information from children under the age of 16. If we learn that we have collected the personal information of a child under 16 we will take steps to delete the information as soon as possible. Please immediately contact us if you become aware that a child under 16 has provided us with personal information.

16. Revisions of this Privacy Policy

This Privacy Policy is valid from 1/1/2026 and replaces any other previous notifications that we had issued in the past regarding our personal data management practices. We reserve the right to revise this Policy at any time. The current version will be always uploaded to our website indicating the date of entry into force, so you know when the most recent revision took place. If there are critical changes in this policy or our personal data practices change significantly in the future, we will notify you by posting the changes on our website.

Consent

By using our website, you hereby consent to our Privacy Policy and agree to its terms.

Annex II – Informed Consent Form

Text in red colour contains guidelines for adjusting this template and should be deleted.

Text in grey colour contains examples and should be adjusted to the context of each activity.

Text included in < > and/or highlighted with yellow should be replaced with content that is suitable to the context of each activity & project as well as to the organisation seeking to obtain the consent.

Before using this template take the time to carefully read and adjust it to the needs of the activity at hand as well as to any relevant regulations and particularities applicable to your country and organisation.

Informed Consent Form

Who we are:

We are < Insert Partner Name > and we are contacting you in the framework of COLOSSUS, a project funded by the European Union under the Horizon Europe Research and Innovation funding programme. A detailed description on how handles personal data is presented in the project's Privacy Policy that accompanies this Consent Form.

Project: COLOSSUS - Multi-purpose unmanned vehicle swarm patrol intelligence platform for port security and resilience (GA Number 101225980).

Partner:

Organisation name: < Insert Partner Name >

Address: < Insert Partner Address >.

Phone: < Insert Partner Phone >.

E-mail: <Insert Partner Generic E-mail Address >

Responsible persons:

You may delete the line referring to the Data Protection Officer if your organisation does not have one.

#	Role	Name	E-mail
1	COLOSSUS Project Manager	<Insert name of project manager from your organisation>	<Insert e-mail of project manager from your organisation >
2	Interviewer	<Insert name of interviewer from your organisation >	<Insert e-mail of interviewer from your organisation >
3	Data Protection Officer	<Insert name of DPO from your organisation >	<Insert e-mail of DPO from your organisation >

What do we need from you?

Please explain in a brief paragraph (4-5 lines) the activity and its purpose under the frame of the project.

Example: We need you to participate in an interview that will be carried out by COLOSSUS with the goal of understanding the current state of play in the European civil security market in regards to port resilience in Europe.

The interview is expected to last for no more than < Insert number of minutes > minutes. We will take written notes and we will be keeping an automated transcript of the interview.

Please adapt the following text to accurately depict the type of personal data to be collected.

To effectively conduct this interview, we need to process some of your personal data:

- Your contact details (full name, email, phone number);
- Some basic demographics (age, gender);
- Your professional info (organization, job position, field of expertise);
- Your opinions on the subject matter.

Why do we need your data & what will we do with them?

We need your data to contact you in order to plan and carry out the aforementioned interview and to resolve any ambiguities, questions and other issues that may arise after and as a result of the interview. We also need to record your data to keep track of the interview process. The project's deliverables that will be derived by the interview will not include your personal data or any other information that could identify you. Your personal data will remain on our written notes (interview's transcript) and/or the sound recording we will make during the interview.

We will share your data with a few other COLOSSUS project partners that are also involved in this task and will participate in the drafting of the relevant deliverables. We are also obliged to grant access to your data to:

- EU officials such as our Project Officer for purposes related to project's evaluation;
- EU agencies and other authorities for project's auditing purposes.

We would also be very happy if you gave us your consent to contact you in the future to ask you to participate in other project's activities (e.g., surveys, interviews, project events, COLOSSUS Security Advisory Board, etc.) and also to inform you about the project's progress (e.g., by sending you a newsletter or similar messages).

How can you withdraw your consent?

You should know that you can withdraw your consent at any time by communicating either on the phone or by email with the responsible persons listed in the previous page. With regards to the informational messages and newsletters you can always opt out by simply clicking the link "Unsubscribe" or something similar included at the end of all the relevant messages.

I hereby give my consent to the processing of my personal data needed for:

*(Please, tick the boxes below to confirm that you give us your consent for the respective subject. Any boxes left unticked mean that **you do not consent to the relevant subject**.)*

#	Consent Subject	Tick box
1	My participation in an interview that will be carried out by COLOSSUS to <insert key objective of the interview >	
2	My participation in future activities of COLOSSUS	
3	Receiving newsletters and messages regarding COLOSSUS activities	

_____	_____	_____
Name of participant	Date	Signature

Annex III – Data Subject Request Form

Text in red colour contains guidelines for adjusting this template and should be deleted.

Text included in < > and/or highlighted with yellow should be replaced with content that is suitable to the context of each activity & project as well as to the organisation seeking to obtain the consent.

COLOSSUS

Data Subject Request form

You may delete the data referring to the Data Protection Officer if your organisation does not have one.

CONTACT

<Insert name of responsible Project Manager>	<Insert name of DPO > (Data Protection Officer)
<Insert email of responsible Project Manager>	<Insert e-mail of DPO >

DATA SUBJECT REQUEST FORM

This form should be used to submit a data subject request under the provisions of the European Union General Data Protection Regulation (GDPR).

Submitter Details

Title:	
Name:	
Address:	

TYPE OF REQUEST

Please select the type of request you are making:

☐	Consent Withdrawal
☐	Access request
☐	Rectification of personal data
☐	Erasure of personal data
☐	Restriction of processing of personal data
☐	Personal data portability request
☐	Objection to processing of personal data
☐	Request regarding automated decision making and profiling

PERSONAL DATA INVOLVED

REQUEST DETAILS

REQUEST REASON/JUSTIFICATION

Name: .....

Signature: .....

Date: .....

***

Once completed, this form should be submitted via e-mail to < Insert contact e-mail of Partner > or posted to:

< Insert Partner Name >

< Insert Partner Address >

Annex III – Record of Processing Activities

Table 13. Record of Processing Activities

No	Project Activity / purpose	Data processing activity	Linked WP(s)	Linked Task (s)	Data subjects	Data source	Data category(-ies)	Responsible partner	Involved partner(s)	Type of involvement	Special category (Art. 9 GDPR)	Lawfulness of processing	Transfer to third countries (non EU-EEA)	Transfer to EU from third countries	Recipients	Comments
1	Market analysis and customer needs assessment	Semi-structured interviews	WP1	Task 1.1	Experts, stakeholders, professionals	Data subject	Contact details, professional role, organisation type, opinions, experience	Q-PLAN	ALL	Collection, analysis, use	No	Art. 6(1)(a) – consent	No	No	–	Data anonymised during analysis and reporting
2	Market analysis and customer needs assessment	Online survey	WP1	Task 1.1	End-users, experts, regulators	Data subject	Professional information, organisational characteristics, opinions, preferences	Q-PLAN	ALL	Collection, analysis, use	No	Art. 6(1)(a) – consent	No	No	–	Survey data anonymised or pseudonymised

No	Project Activity / purpose	Data processing activity	Linked WP(s)	Linked Task (s)	Data subjects	Data source	Data category(-ies)	Responsible partner	Involved partner(s)	Type of involvement	Special category (Art. 9 GDPR)	Lawfulness of processing	Transfer to third countries (non EU-EEA)	Transfer to EU from third countries	Recipients	Comments
3	Regulatory, ethics and security requirements analysis	Desk research and expert inputs	WP1	Task 1.2	–	Public sources	Policy documents, standards, regulatory texts	iNGS	ALL	Collection, analysis, use	No	Not applicable	No	No	–	No personal data processed
4	System requirements co-design	Requirements workshops and analysis	WP1	Task 1.3	Consortium experts	Data subject	Professional opinions, technical inputs	UNINOVA	ALL	Collection, use	No	Art. 6(1)(f) – legitimate interest	No	No	–	Internal consortium processing
5	System development and integration	USV technical development and	WP2	Task 2.1	–	System-generated	Sensor data, telemetry, logs, performa	TV	UNINOVA, iNGS, AASA	Generation, processing, use	No	N/A	No	No	–	Non-personal technical data

No	Project Activity / purpose	Data processing activity	Linked WP(s)	Linked Task (s)	Data subjects	Data source	Data category(-ies)	Responsible partner	Involved partner(s)	Type of involvement	Special category (Art. 9 GDPR)	Lawfulness of processing	Transfer to third countries (non EU-EEA)	Transfer to EU from third countries	Recipients	Comments
		testing					nce metrics									
6	System development and integration	UAV & UGV technical development	WP2	Task 2.2	–	System-generated	Image data, depth data, LiDAR point clouds, navigation data	UNINOVA	iNGS, AASA, TV	Generation, processing, use	No	N/A	No	No	–	Non-personal technical data
7	Secure data exchange	Secure interfaces and communications testing	WP2	Task 2.3	–	System-generated	Network logs, encryption metadata, interface logs	iNGS	UNINOVA, AASA, TV	Generation, processing, use	No	N/A	No	No	–	No personal data processed
8	AI and data fusion	Multi-modal data	WP3	Task 3.1	–	System-gen	Fused sensor data,	UNINOVA	iNGS, TV	Processing, use	No	N/A	No	No	–	Internal technical datasets

No	Project Activity / purpose	Data processing activity	Linked WP(s)	Linked Task (s)	Data subjects	Data source	Data category(-ies)	Responsible partner	Involved partner(s)	Type of involvement	Special category (Art. 9 GDPR)	Lawfulness of processing	Transfer to third countries (non EU-EEA)	Transfer to EU from third countries	Recipients	Comments
		harmonisation				erated	harmonised datasets									
9	AI-based autonomy	Intelligent swarm navigation	WP3	Task 3.2	–	System-generated	Coordination messages, state data, mission logs	iNGS	UNINOVA, TV	Processing, use	No	N/A	No	No	–	No personal data
10	Underwater inspection analytics	3D reconstruction and modelling	WP3	Task 3.3	–	System-generated	Sonar data, visual data, 3D models	UNINOVA	TV, iNGS	Processing, use	No	N/A	No	No	–	No personal data
11	System validation	Integration lab testing and fine tuning	WP4	Task 4.1	–	System-generated	Configuration files, logs, test results	TV	UNINOVA, iNGS, AASA	Processing, use	No	N/A	No	No	–	Technical validation data

No	Project Activity / purpose	Data processing activity	Linked WP(s)	Linked Task (s)	Data subjects	Data source	Data category(-ies)	Responsible partner	Involved partner(s)	Type of involvement	Special category (Art. 9 GDPR)	Lawfulness of processing	Transfer to third countries (non EU-EEA)	Transfer to EU from third countries	Recipients	Comments
12	System validation	Stress-testing and technology validation	WP4	Task 4.2	–	System-generated	Performance metrics, failure logs	TV	UNINOVA, iNGS, AASA	Processing, use	No	N/A	No	No	–	Non-personal data
13	System supervision	COLOSSUS Web Platform operation	WP4	Task 4.3	Operators (limited)	System-generated / data subject	User interaction logs, alerts, system status	TV	ALL	Processing, use	No	Art. 6(1)(f) – legitimate interest	No	No	–	Access-controlled, non personal data
14	Pilot preparation and deployment	Operational pilot trials	WP5	Task 5.1	Operators, staff	Data subject / system-gen	Operational logs, performance data, observations	AASA	ALL	Collection, use	No	Art. 6(1)(a) – consent	No	No	–	Feedback anonymised

No	Project Activity / purpose	Data processing activity	Linked WP(s)	Linked Task (s)	Data subjects	Data source	Data category(-ies)	Responsible partner	Involved partner(s)	Type of involvement	Special category (Art. 9 GDPR)	Lawfulness of processing	Transfer to third countries (non EU-EEA)	Transfer to EU from third countries	Recipients	Comments
						erated										
15	Technical evaluation and market validation	Evaluation and assessment activities	WP5	Task 5.2	Experts, evaluators	Data subject	Professional feedback, evaluation results	AASA	Q-PLAN, TV	Collection, analysis, use	No	Art. 6(1)(a) – consent	No	No	–	Aggregated reporting
16	Replication and policy support	Replication guidelines & policy recommendations	WP5	Tasks 5.3–5.4	Experts, policy makers	Data subject	Professional opinions, policy inputs	AASA	UNINOVA, Q-PLAN	Collection, use	No	Art. 6(1)(a) – consent	No	No	–	No attribution
17	Dissemination and communication	Events, social media, newsletters	WP6 – WP7	Tasks 6.1 / 7.1	Subscribers, participants	Data subject	Contact details, engagement metrics	Q-PLAN	ALL	Collection, use	No	Art. 6(1)(a) – consent	No	No	–	Opt-in, unsubscribe available

No	Project Activity / purpose	Data processing activity	Linked WP(s)	Linked Task (s)	Data subjects	Data source	Data category(-ies)	Responsible partner	Involved partner(s)	Type of involvement	Special category (Art. 9 GDPR)	Lawfulness of processing	Transfer to third countries (non EU-EEA)	Transfer to EU from third countries	Recipients	Comments
18	Exploitation and collaboration	Business development and synergies	WP6 – WP7	Tasks 6.3 / 7.5	Stakeholders	Data subject	Professional contact details	Q-PLAN	ALL	Collection, use	No	Art. 6(1)(f) – legitimate interest	No	No	–	Minimal data use
19	Governance and ethics	Security Advisory Board operation	WP8 – WP9	Tasks 8.2 / 9.2	SAB members	Data subject	Contact details, professional role, opinions	UNINOVA	ALL	Collection, use	No	Art. 6(1)(a) – consent	No	No	–	Internal governance use

GA 101225980

PARTNERS

FOLLOW US

COLOSSUS project @colossus_eu @colossus_eu

Co-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 REA. Neither the European Union nor the granting authority can be held responsible for them.

Co-funded by the European Union