



ENVRI-Hub NEXT

D2.2

Update on the DMP

Status: Under EC Review

Dissemination Level: Public



Funded by
the European Union

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

Abstract**Key Words**

Data Management, FAIR, GDPR, Metadata, Re-usable Data

This updated Data Management Plan (DMP) outlines how research data in the ENVRI-Hub NEXT project continues to be collected, processed, monitored, and catalogued according to the FAIR principles. Building on the initial DMP (D1.2 – Data Management Plan), this deliverable refines existing workflows and incorporates new data practices introduced as the project progresses. As a living document, it evolves alongside the project to ensure that data-management procedures remain accurate, compliant, and aligned with emerging technical and organisational needs.

Revision History

Version	Date	Description	Author/Reviewer
V 0.1	11/06/2026	ToC and First Draft	Matteo Agat (EGI)
V 0.2	14/07/2026	Draft ready for Review	Matteo Agat (EGI)
V 0.3	24/07/2026	External Review	Hien Bui (EGI) Marta Gutierrez (EGI)
V 1	31/07/2026	Final Version	Matteo Agat (EGI)

Document Description			
D2.2 – Update on the DMP			
Work Package Number 2			
Document Type	Deliverable		
Document Status	Under EC Review	Version	1
Dissemination Level	Public		
Copyright Status	 This material by Parties of the ENVRI-Hub NEXT Consortium is licensed under a Creative Commons Attribution 4.0 International License.		
AI Declaration	The authors acknowledge the use of GenAI chatbots for non-substantial use, e.g. improving the readability and language of the text, as a basic support tool. [Gemini 3.6 Flash]		
Lead partner	EGI Foundation		
Document Link	https://documents.egi.eu/document/4028		
DOI	https://zenodo.org/records/21722936		
Author(s)	Matteo Agati (EGI)		
Reviewers	• Hien Bui (EGI) • Marta Gutierrez (EGI)		
Moderated by:	Matteo Agati (EGI)		
Approved by:	Development Steering Board (DSB)		

Terminology / Acronyms	
Term/Acronym	Definition
API	Application Programming Interface
DMP	Data Management Plan
DocDB	Document Database
DOI	Document Object Identifier
DPO	Data Protection Officer
ECV	Essential Climate Variables
EOSC	European Open Science Cloud
ENVRI	Environmental Research Infrastructures
FAIR	Findable, Accessible, Interoperable, and Reusable
FDP	FAIR Data Point
GDPR	EU General Data Protection Regulation
LLM	Large Language Model
ORE	Open Research Europe
RI	Research Infrastructure
SSO	Single Sign On
URL	Uniform Resource Locator

Table of Contents

Executive Summary.....	7
1. Introduction.....	8
1.1. Purpose and Scope of the Document.....	8
1.2. Structure of the Document.....	8
2. Data Summary.....	10
2.1. Potential Data Assets Generated by the Project.....	10
2.2. Types of Data.....	10
2.3. Existing Data/Software and Newly Generated Project Outputs.....	10
3. FAIR Data.....	12
3.1. Making Data Findable, including Provisions of Metadata.....	12
3.1.1. Metadata.....	12
3.2. Making Data Openly Accessible.....	14
3.2.1. Repositories.....	14
3.2.2. Standardised Access Protocol.....	14
3.2.3. Metadata Availability.....	14
3.3. Making Data Interoperable.....	15
3.4. Increase Data Re-use.....	15
4. Allocation of Resources.....	17
4.1. Data Management Responsibilities.....	17
5. Data Security.....	18
6. Service-Level Data Protection and Access Policies.....	19
7. Conclusion and Next Steps.....	21
8. Appendix I: RIs Datasets Assessment.....	22

Table of Tables

- [Table 1 - Repository Metadata](#)
- [Table 2 - Document Metadata](#)
- [Table 3 - Policy Framework for ENVRI-Hub Services](#)

Executive Summary

This updated ENVRI-Hub NEXT Data Management Plan (DMP) refines the project's strategy for managing research data throughout its lifecycle. It builds on the initial DMP by reaffirming the application of the FAIR principles (Findable, Accessible, Interoperable, and Reusable), promoting transparency, accessibility, and reusability. It also incorporates updated practices, tools, and workflows introduced during the project, ensuring that data-management procedures remain coherent and aligned with the project's evolving needs.

Key points of the deliverable include:

- **Purpose and Scope:** The DMP aims to standardise data management practices across the project, ensuring that all data is collected, processed, and stored in a manner that adheres to the highest standards of scientific integrity and openness.
- **Data Management Strategies:** Detailed plans for making data findable, accessible, interoperable, and reusable are presented. This includes the use of established metadata standards, open-access repositories, and interoperable data formats.
- **Resource Allocation:** Clear delineation of roles and responsibilities related to data management, ensuring that all project partners understand their obligations and contributions.
- **Data Security and Ethical Considerations:** Robust measures are in place to protect data integrity and security, along with comprehensive ethical guidelines to address any potential issues related to data use and sharing.
- **Sustainability and Long-term Preservation:** Strategies for the long-term preservation of data, including the use of trusted repositories and adherence to open-access policies, ensuring that data remains accessible and usable beyond the project's duration.

The deliverable concludes with recommendations to support strong and consistent data-management practices across the consortium. As the final version of the ENVRI-Hub NEXT Data Management Plan, it consolidates all procedures, standards, and FAIR-aligned workflows that will guide data handling for the remainder of the project. By establishing a clear and stable framework, the project reinforces its commitment to high-quality data stewardship and sets a benchmark for data management in environmental research, fostering openness, accessibility, and collaboration within the community.

1. Introduction

1.1. Purpose and Scope of the Document

The purpose of this document is to present the updated Data Management Plan (DMP) for the ENVRI-Hub NEXT project. It outlines the principles, processes, and responsibilities governing how project data are collected, processed, stored, shared, and made FAIR throughout the project lifecycle. As the final version of the DMP, it consolidates all established workflows and standards, ensuring that data-management practices remain coherent and consistently applied across the consortium. The document also clarifies the boundaries of the DMP by specifying the types of data covered, the actors involved, and the procedures that support long-term accessibility and reusability.

Compared to the previous version of the Data Management Plan¹, the most significant change is the replacement of Section 6, previously titled "*Ethics Management*". In the earlier DMP, this section provided a high-level overview of ethics procedures; however, the project now has three dedicated ethics deliverables, the most recent submitted in July 2025 (*D15.2 OEI - Requirement No. 2*), which comprehensively report all ethics-related processes, assessments, and compliance measures. As a result, [Section 6](#) has been restructured and renamed *Service-Level Data Protection and Access Policies*. This new section focuses on the policy instruments governing the ENVRI-Hub services running on the EGI Cloud, summarising the Privacy Notices, Access and Use Policies, and Cookies Policies that ensure transparent, compliant, and responsible handling of user and service-level data.

In addition, [Appendix I](#) includes the full Data Management Plan included in the above-mentioned last ethics deliverable. This appendix will be updated again at the end of the project with the delivery of *D15.3 OEI - Requirement No. 3* to reflect final data-management practices and ensure continuity and alignment between the ethics reporting and the DMP.

Through this comprehensive scope, the DMP provides a stable reference for all partners and ensures alignment with European best practices for research data stewardship.

1.2. Structure of the Document

This document comprises the following chapters:

- [Section 1](#) presents an introduction to the document.
- [Section 2](#) presents the purpose of data collection, type and format, and origin of the data.
- [Section 3](#) outlines the ENVRI-Hub NEXT FAIR data strategies.
- [Section 4](#) briefly describes the allocation of resources.
- [Section 5](#) summarises the project's core data-security and preservation measures.
- [Section 6](#) outlines the policy framework governing all ENVRI-Hub services hosted on the EGI Cloud, summarising the key documents that regulate data protection, user access, and web-based interactions across the platform.

¹ <https://zenodo.org/records/18611947>

- [Section 7](#) provides the conclusions of this deliverable.
- [Appendix I](#) includes the assessments carried out by the ENVRI Research Infrastructures on the environmental datasets used in the project, originally compiled and reported in Deliverable D15.2 OEI – Requirement No. 2 (submitted in July 2025).

As this part serves as a reference framework and a set of guiding principles for the project duration, the original present tense and modal formulations (e.g. “shall”) have been preserved intentionally.

2. Data Summary

This section provides an overview of the potential data assets generated by the ENVRI-Hub NEXT project, including the types and origins of data involved. The data generated will adhere to the FAIR principles, ensuring it is findable, accessible, interoperable, and reusable.

2.1. Potential Data Assets Generated by the Project

The ENVRI-Hub NEXT project generated and will generate a wide range of data assets, including:

- **Research Data:** processed data from environmental research activities.
- **Software:** ENVRI-Hub software, tools, software services, scripts, workflows and computational notebooks developed for data analysis, processing, and visualisation.
- **Documentation:** manuals, guidelines, and training materials to support data use and software applications.
- **Metadata:** descriptive information about the data assets to ensure discoverability and proper usage.
- **Publications:** research papers and reports detailing project findings and methodologies.

2.2. Types of Data

The project will manage various types of data, including but not limited to:

- **Textual Data:** documents, reports, and publications in formats such as .docx, .pdf, .txt, and .rtf.
- **Numerical Data:** datasets in formats like .csv, .xls/.xlsx, and .netcdf.
- **Code and Software:** programming scripts, software packages and APIs in formats such as .py, .R, .sh, .c/c++, .ipynb, .java and .js.
- **Multimedia Data:** images, videos, and audio files in formats like .jpg/.jpeg, .png, .gif, .mp4, and .wav.
- **Metadata Records:** information about the data in formats like .xml, .json, .rdf, and .ttl.
- **Presentations:** slides and visual aids in formats such as .ppt/.pptx and .pdf.

These formats ensure that the project outputs are accessible, interoperable, and reusable, adhering to the FAIR principles. The selection of specific formats is guided by community standards and the nature of the data.

2.3. Existing Data/Software and Newly Generated Project Outputs

The project builds on both existing and newly generated data and software outputs:

Existing Data/Software

ENVRI-HUB-NEXT is not responsible for external data/software. Thus, this DMP does not cover external datasets or software listed in the following bullet points. Here, individual DMPs issued by RIs and/or Projects like ENVRI FAIR are in place.

- **Content in the current ENVRI-Hub:** the pre-existing catalogue and knowledge base include datasets, software, and services that have been curated according to FAIR principles.
- **Research Infrastructure Databases:** databases from RIs such as ICOS ERIC², LifeWatch ERIC³, and EPOS ERIC⁴, among others, provide a rich source of reusable data.
- **Software:** existing ENVRI-Hub prototype and software tools for data processing, analysis, and visualisation, such as those hosted on GitLab, will be utilised and extended.
- **Metadata Standards:** established metadata standards and schemas from prior projects and RIs.

In contrast to the external inputs listed above, the project is fully responsible for the creation, management, and long-term stewardship of all newly generated outputs. These items fall within the direct scope of this DMP, as they originate from ENVRI-Hub NEXT activities and must therefore comply with the project's FAIR-aligned data-management procedures and quality standards.

Newly Generated Project Outputs

- **New Datasets:** generated from ongoing research activities within the project, these datasets will be documented and published following the FAIR principles.
- **Enhanced/New Software:** development of new research features of ENVRI-Hub, software tools, and significant enhancements to existing tools will support advanced data processing, interoperability, and visualisation needs.
- **Metadata Records:** creation of comprehensive metadata records for all new datasets and software outputs will ensure discoverability and reusability.
- **Documentation and Training Materials:** new documentation and training materials (including FitSM⁵ training) will support the use and understanding of the data and tools generated by the project.
- **Publications:** research papers and project deliverables will detail the methodologies, findings, and advancements achieved through the project activities.

² <https://www.icos-cp.eu/>

³ <https://www.lifewatch.eu/>

⁴ <https://www.epos-eu.org/epos-eric>

⁵ <https://www.fitsm.eu/>

3. FAIR Data

The data generated will be managed according to FAIR principles in line with Open Science principles. To ensure that data and project outputs are easily discoverable, they will be made accessible through the ENVRI-Hub NEXT Community in Zenodo⁶ and other reliable open repositories. Access will be provided according to open-access policies, and interoperability will be promoted by including all essential information, data, and metadata along with the reports.

Lastly, to ensure reusability, data and project outputs will be shared under permissive licenses, such as CC BY, whenever feasible.

3.1. Making Data Findable, including Provisions of Metadata

Data will be designed to be FAIR and will be made available as openly and as early as possible. Data will only be restricted when required by regulatory or legal constraints or due to the owner's legitimate interests. Raw data from consultations (like user surveys and questionnaires) will be aggregated and anonymised to ensure it can be reused for future purposes.

3.1.1. Metadata

The metadata required by data repositories will be used as outlined in [Table 1](#). For documents (such as project deliverables and presentations), ENVRI-Hub NEXT applies a standard metadata schema to ensure consistent documentation of all assets, as shown in [Table 2](#). All documents (Public and Confidential) are stored in the EGI Document Database (DocDB)⁷, while Public documents are additionally deposited in the ENVRI-Hub NEXT Zenodo community for open access and long-term preservation.

Table 1 – Repository Metadata

Element	Definition
Abstract	An abstract describing the document's contents and main conclusions. On submission of the final version, this should be entered into the relevant field in the repository metadata.
Keywords	A list of relevant terms that help describe the content of the deliverable and support search and categorisation into the Zenodo portal.
Title	The title of the document, datasets, etc.
Submitter	The person submitting the document to the repository.
Authors	Individuals who structure and write an important part of the document's content.
DOI	The URL to the Zenodo (or other) document repository.
Publication date	The date of the publication.

⁶ <https://zenodo.org/communities/envrihubnext/>

⁷ <https://documents.egi.eu/>

Version	The version number generated by the document repository for the repository identifier. Versioning rule: · +0.1 – a new version of the draft · +1.0 – a new version of the approved document
Language	The language of the intellectual content of the resource.
Communities	A specific community in which the upload will appear.
License	The legal standing of the document in terms of copyright protection.
Modify	The groups that can modify the document.

Table 2 - Document Metadata

Ele	Definition
Abstract	An abstract describing the document's contents and main conclusions. On submission of the final version, this should be entered into the relevant field in the repository metadata.
Keywords	A list of relevant terms that help describe the content of the document and support search and categorisation into the Zenodo portal.
Title	Title of the document as described in the Description of Action (in case of Deliverables and Milestones).
WP	The number of the Work Package to which the document is associated.
Copyright status	Public material is licensed under a Creative Commons Attribution 4.0 International License Confidential material may not be reproduced, distributed, or disclosed in whole or in part without the prior written consent of the ENVRI-Hub NEXT consortium.
Document Type	The type of the document (e.g., Report, Demonstrator, White Paper...);
Status	• Draft - The document is being prepared. • Under EC review - The document is submitted to the EC portal and has not yet been approved by the European Commission. • Approved by EC - The document is approved by the European Commission. • Final - Status of the document uploaded to the EC portal.
Dissemination Level	• Public - Can be shared without restrictions. • Confidential - Can be shared only with the European Commission and project partners.
Document link	The URL in the document repository that provides access to the document on DocDB.
Digital Object Identifier (DOI)	The URL to the Zenodo (or other) document repository
Authors	Individuals who structure and write an important part of the document's content.
Reviewers	Subject matter experts and stakeholders who critically evaluate the deliverable to verify its accuracy, relevance, and overall quality

Approval

The Project Board is responsible for formally approving the document

3.2. Making Data Openly Accessible

Textual outputs such as reports and presentations, except for administrative documents containing sensitive legal, personal, or financial information, will be made available to the ENVRI-Hub NEXT Community on Zenodo upon release. These outputs will be shared under a permissive licence, such as CC BY, and accompanied by comprehensive metadata under CC05. Tools and code will be extensively documented and preserved through open repositories like GitLab and provided using licences like GPL-3⁸ or Apache-2⁹.

The metadata will include details like authorship, ownership, provenance, access rights, and persistent identifiers (PIDs), all in alignment with the FAIR principles.

3.2.1. Repositories

All documents, presentations and other materials that form an official output of the project (not just milestones and deliverables) are placed in the EGI's own trusted document repository¹⁰ to provide a managed central location for all materials.

In addition, public deliverables and publications will be shared publicly via the Zenodo platform to increase the discoverability of the project outputs.

All profiles, specifications, configuration files, software, workflows, and code will be deposited in Zenodo. Therefore, ENVRI-Hub NEXT will use DocDB, Zenodo, and GitLab as its standard and main repositories.

3.2.2. Standardised Access Protocol

All data will be accessible via a uniform resource locator (URL) or Document Object Identifier (DOI). There will be no restrictions on the use of the research outputs, both during and after the end of this project.

The project frequently interacts with the ENVRI Board¹¹, composed of RI representatives, to ensure that outputs are all-inclusive to the ENVRI Community. It will also take into consideration public engagement from end users¹² and citizens through interactive workshops at targeted events, consultations and social media.

3.2.3. Metadata Availability

Metadata containing information to enable users to access the data will be openly available and published via the ENVRI-Hub Catalogue of Services¹³ and Catalogue of Data¹⁴ together with the

⁸ <https://www.gnu.org/licenses/gpl-3.0.en.html>

⁹ <https://www.apache.org/licenses/LICENSE-2.0>

¹⁰ <https://documents.egi.eu/>

¹¹ <https://envri.eu/governance/>

¹² <https://envri.eu/envri-hub-user-group/>

¹³ <https://catalogue.envri.euthe/>

¹⁴ [https://fdp.envri.euthe/responsibility/Infrastructures/responsible for Infrastructures/](https://fdp.envri.euthe/responsibility/Infrastructures/responsible%20for%20Infrastructures/)

data (in this case, Research Infrastructures are responsible for data availability), in the same repositories as listed under [Section 3.2.1](#). There is no time limit on metadata and data availability.

The ENVRI-Hub NEXT project acknowledges the value of documentation for interoperability purposes, increases uptake by different communities, and encourages data owners to document their research data assets. ENVRI-Hub NEXT does not enforce specific provisions on documentation as long as the data asset is hosted on one of the mentioned repositories and properly curated according to the repository's best practices.

Research data (published via FAIR Data Points¹⁵, Catalogue of Services¹⁶ or Knowledge Base¹⁷ with metadata standards DCAT-AP¹⁸ and contextualised LLMs for further analysis) itself should not be considered self-documenting, and each published asset must be associated with sufficient documentation resources accessible through a public URL. Documentation must be browsable and include hypertext references to facilitate its fruition. Recommended documentation formats include Markdown, HTML, and other markup languages. The inclusion of machine-readable documentation, such as OpenAPI, where applicable, is thoroughly encouraged. If a scientific publication is tied to the research data asset, the publication itself should be referenced and/or made available as part of the documentation.

3.3. Making Data Interoperable

The ENVRI-Hub NEXT consortium recognises that data interoperability is essential for ensuring that published data assets can be meaningfully compared, merged, and integrated across Research Infrastructures. To support this, the consortium adopts and promotes established interoperability standards, including DCAT-AP for catalogues, OpenAPI specifications for the VRE Library, REST APIs for machine-to-machine service communication, and iAdopt for harmonising scientific variables. Interoperability is further strengthened through ENVRI-ID, which provides unified access to all participating RIs and relies on widely accepted authentication and authorisation standards such as OIDC Connect and SAML. Through these measures, the consortium is committed to enforcing robust, standards-based interoperability across the entire ENVRI-Hub ecosystem.

3.4. Increase Data Re-use

The project implements measures to increase data reuse by providing clear documentation, adhering to FAIR principles, and ensuring data is accessible and interoperable. Regular training and documentation of services, metadata or analytical workflows will be provided and made available to the training platform, assisted by training events to maximise the reusability of the data.

Peer-reviewed research papers produced during the project will adhere to the European Commission's Open Access policy for publications. Preprints will be deposited in reputable open repositories, such as arXiv or institutional archives, whenever feasible. Upon publication, these

¹⁵ <https://fdp.envri.eu/>

¹⁶ <https://catalogue.envri.eu/>

¹⁷ <https://search.envri.eu/se/arch>

¹⁸ <https://op.europa.eu/en/web/eu-vocabularies/dcat-ap>

papers will be self-archived in Zenodo. Ideally, they will also be published in open-access journals or platforms, with Open Research Europe (ORE) as the preferred choice, or in selected Diamond Open Access journals. All articles will be issued with open-access licences that impose minimal restrictions. Both the papers and related materials will be made available in the ENVRI-Hub NEXT Community on Zenodo as needed.

4. Allocation of Resources

Any expenses associated with the collection/production of FAIR data during the ENVRI-Hub NEXT activities are included in the project budget. These expenditures will be required to cover a variety of particular data processing and data management operations, ranging from data collection and documentation to storage and preservation to distribution and re-utilisation.

These operations are a component of the Work Packages (WPs) that process the relevant data; hence, the needed effort will be part of the relevant WP.

The expenses of long-term data preservation are minimal when using the EGI Online Storage and Google Drive platforms. Using Zenodo and GitLab (both free of charge) ensures that costs for long-term preservation of the data are manageable.

4.1. Data Management Responsibilities

Within the ENVRI-Hub NEXT project, the following roles and responsibilities are associated with Data Management, which are defined as follows:

- **WP leaders** are in charge of organising the data processing and quality assurance that takes place inside the Work Package they are leading.
- **Task Leaders** are in charge of the data compiled/produced throughout the operation of the task that they are responsible for. In addition to that, they also make sure that the data are properly prepared to be shared among the partners and made publicly available, when applicable.
- **Data users** are consortium partners who use, e.g., processing operations on the compiled/produced data.
- **Quality and Risk Manager** monitors and supports the WP leaders and Task Leaders/Use Case leaders in keeping the DMP Confluence pages up to date. In addition, report the changes and processes via milestones and deliverables as specified in the Grant Agreement.

5. Data Security

Any gathered data will be securely handled throughout the entire duration of the ENVRI-Hub NEXT project to protect it from loss and unauthorised access. Personal data is only accessible to those who are authorised¹⁹ to access it.

All partners/beneficiaries responsible for processing personal data²⁰ have the responsibility to ensure that the data remains protected under all necessary security controls (including backup policies and integrity checks²¹) and access controls (identification, authentication, authorisation) within their infrastructure. In the unfortunate event of a personal data breach, the project partners will notify without delay their competent national supervisory authorities as well as the data subject(s) that may be affected by the breach. At the same time, they will document any personal data breaches and all related information.

Regarding open data, for security and long-term preservation, ENVRI-Hub NEXT relies on the EGI Document Repository, Zenodo, Google Drive and GitLab platforms. The ENVRI-Hub services operate CI/CD pipelines that ensure Quality Assurance and check for software vulnerabilities. A security policy is in place to ensure that incident response is handled as described in FitSM processes.

¹⁹ Personal contact data are collected during the [EGLSSO](#) (Single Sign On) account creation by the users. After the access is granted, the users can manage their data in autonomy.

²⁰ Processing, according to Regulation (EU) 2017/679 of the European Parliament (GDPR), means any operation or set of operations which is performed on personal data or on sets of personal data, whether 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

²¹ The integrity check is the process of comparing the current state of stored data and/or programs to a previously recorded state to determine any alteration or change.

6. Service-Level Data Protection and Access Policies

As part of the ENVRI-Hub service ecosystem deployed on the EGI Cloud, each platform component operates under clearly defined policy instruments that ensure responsible data handling, transparent user interaction, and compliance with European legal and ethical standards. To support clarity and accountability, the project documents for each service its Privacy Notice, Access and Use Policy, Cookies Policy, and Terms and Conditions, which together describe how user data is processed, how the service may be accessed and utilised, and how web-based interactions are managed ([Table 3](#)).

These policy elements are directly relevant to the Data Management Plan, as they define the conditions under which users interact with the ENVRI-Hub services, clarify how personal and usage data are handled, and demonstrate compliance with GDPR and other applicable regulations. Including them ensures transparency across the service landscape and shows that each component of ENVRI-Hub – ENVRI-ID, Catalogue of Services (CoS), ENVRI Catalogue of Data (CoD), ENVRI Core Services (Core), ENVRI Innovation Hub (EIH), and the Knowledge Base (KB) – operates with clear, documented safeguards that support responsible data governance.

Table 3 - Policy Framework for ENVRI-Hub Services

ENVRI-ID			
Documents	Role	Organisation	Publicly accessible link
Privacy Notice	Data Controller	EGI	Link
	Data Processor	GRNET	
Access and Use Policy	n/a	n/a	Link
Cookies Policy	n/a	n/a	Link
Catalogue of Services			
Privacy Notice	Data Controller	EGI	Link
	Data Processor	GRNET	
Access and Use Policy	n/a	n/a	Link
Cookies Policy	n/a	n/a	Link
ENVRI Catalogue of Data			
Privacy Notice	Data Controller	ICOS ERIC	Link
	Data Processor	EGI	
Terms and Conditions	n/a	n/a	Link
ENVRI Core Services (Core)			
Privacy Notice	Data Controller	FZJ	Link

	Data Processor	EGI	
Terms and Conditions	n/a	n/a	Link
ENVRI Innovation HUB			
Privacy Notice	Data Controller	FZJ	Link
	Data Processor	EGI	
Terms and Conditions	n/a	n/a	Link
Knowledge Base			
Privacy Notice	Data Controller	FZJ	Link
	Data Processor	EGI	
Terms and Conditions	n/a	n/a	Link

7. Conclusion and Next Steps

ENVRI-Hub NEXT DMP is a complete data management approach that complies with Horizon Europe recommendations and aims to make data as findable, accessible, interoperable, and reusable (FAIR) as feasible.

Relying on robust technological solutions and standards, such as the EGI Document Repository, Zenodo, GitLab and Google Drive²² for the execution of these processes ensures a secure, consistent, and sustainable approach to data handling. Additionally, this will ensure that the data created or compiled throughout the ENVRI-Hub NEXT project, including open data and public publications, will be kept and continue to be usable once the project is completed.

The plan is intended to safeguard the analysis of compiled/created data based on the privacy level and to use an alternate sharing methodology relying upon this level. Confidential information or information that raises ethical problems will not be released.

Tasks 1.3/2.3 implement the DMP with the help of data managers of the RIs under the supervision of the coordinator. All RIs have ample experience in developing DMPs and FAIR management of research data in research projects and their own RI, as well as with curation and preservation issues, as the reason for the existence of the RI data centres involved in the project.

Finally, the DMP is built on guaranteeing appropriately informed consent and protecting each participant's zone of privacy, while adhering to GDPR guidelines.

²² Google is listed in the DPF and thus, covered by EU-U.S. Data Privacy Framework <https://www.dataprivacyframework.gov/list>. All instances of Jira, GITLAB, Confluence are operated on European Servers and compliant with GDPR.

8. Appendix I: RIs Datasets Assessment

In this Appendix, we present the assessment carried out by the ENVRI Research Infrastructures on the environmental datasets used in the project, originally compiled and reported in Deliverable D15.2 *OEI – Requirement No. 2* (submitted in July 2025).

The assessment of the environmental datasets was conducted under the ENVRI RIs' responsibilities (ACTRIS²³, AnaEE ERIC²⁴, EPOS ERIC²⁵, Euro-Argo ERIC²⁶, IAGOS AISBL²⁷, ICOS ERIC²⁸, LifeWatch ERIC²⁹, eLTER³⁰, SeaDataNet³¹), and outlines the project's approach for creating, re-using, and managing research data in order to align with FAIR principles and legal obligations. It describes what types of data are expected, their size, source, formats, and their relation to the project's goals. Attention is also paid to the long-term value of data beyond the immediate research, including utility for other stakeholders.

The assessment covered practices for ensuring data is findable, accessible, interoperable, and reusable. It addresses metadata standards, repository use, licensing, and restrictions on access. Data security and informed consent are also examined, especially regarding personal or sensitive information. Finally, we reflect on the responsible handling of related outputs and the broader ethical context in which the data lifecycle operates.

²³ <https://www.actris.eu/>

²⁴ <https://www.anaee.eu/>

²⁵ <https://www.epos-eu.org/epos-eric>

²⁶ <https://www.euro-argo.eu/>

²⁷ <https://www.iagos.org/>

²⁸ <https://www.icos-cp.eu/>

²⁹ <https://www.lifewatch.eu/>

³⁰ <https://elter-ri.eu/>

³¹ <https://www.seadatanet.org/>

ACTRIS

RI Name	ACTRIS	
Contact	Markus Fiebig (mf@nilu.no)	
Data Summary (Optional)		
Will you re-use any existing data and what will you re-use it for?	ACTRIS re-uses its own and other providers' data to generate combined data products (Data analysis products produced from one or several primary observation and/or model data products)	
Will you re-use any existing data and will this generate new data?	This generates new data, which is made accessible to the public according to the involved licences.	
What types and formats of data will the project generate or re-use?	Data products include scalar time series, profile time series, and time series of fields, which are usually served in NetCDF format.	
What is the purpose of the data generation or re-use and its relation to the objectives of the project?	Combined products are supposed to make the data more accessible and usable for the primary user groups and use cases.	
What is the expected size of the data that you intend to generate or re-use?	Varies largely depending on the data type between GB and TB orders.	
What is the origin/provenance of the data, either generated or re-used?	Produced by RI internal observations and external data providers.	
To whom might your data be useful ('data utility'), outside your project?	Primary user groups outside the RI include operational atmospheric services, e.g. Copernicus,	
FAIR Data		
1.) Making data findable, including provisions for metadata	Will data be identified by a persistent identifier?	Yes
	Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.	Yes. ACTRIS has defined its own internal metadata schema, which is designed to map to major domain standards.
	Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?	ACTRIS metadata uses, wherever possible, controlled vocabulary from its own, FAIR-compliant,

		I-ADOPT-compliant, machine-readable vocabulary
	<i>Will metadata be offered in such a way that it can be harvested and indexed?</i>	Yes
2.) Making data openly accessible		
a) Repository:	<i>Will the data be deposited in a trusted repository?</i>	Yes
	<i>Have you explored appropriate arrangements with the identified repository where your data will be deposited?</i>	Not applicable, ACTRIS DC is trusted.
	<i>Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?</i>	2 x Yes
b) Data:	<i>Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.</i>	Yes
	<i>If an embargo is applied to give time to publish or seek the protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	Not applicable
	<i>Will the data be accessible through a free and standardised access protocol?</i>	Yes
	<i>If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?</i>	No restrictions besides the requirement to attribute.
	<i>How will the identity of the person accessing the data be ascertained?</i>	Not necessary, since the data aren't access restricted.
	<i>Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?</i>	No

c) Metadata:	<i>Will metadata be made openly available and licensed under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?</i>	Yes, but licensed as CC-BY 4.
	<i>How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?</i>	For the lifetime of ACTRIS RI.
	<i>Will documentation or reference about any software be needed to access or read or process the data be included? Will it be possible to include the relevant software (e.g. in open-source code)?</i>	No.
3.) Making data interoperable	<i>What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?</i>	Data will be in NetCDF format with CF / ACDD profile, using ACTRIS vocabulary.
	<i>In case it is unavoidable that you use uncommon or generate project-specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining, or extending them?</i>	Yes.
	<i>Will your data include qualified references to other data (e.g. other data from your project, or datasets from previous research)?</i>	Yes.
4.) Increase data re-use	<i>How will you provide the documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?</i>	Metadata is included with the data.
	<i>Will your data be made openly available in the public domain to permit the widest re-use possible? Will your data be licensed using standard re-use licenses, in line with the obligations set out in the Grant Agreement? Under which license?</i>	Yes, using CC BY 4 licence.

	Will the data produced in the project be useable by third parties, in particular after the end of the project?	Yes.
	Will the provenance of the data be thoroughly documented using the appropriate standards?	In progress, PROV-O will be used.
	Describe all relevant data quality assurance processes.	
Other research outputs		
In addition to the management of data, are you also considering and planning for the management of other research outputs that may be generated or re-used throughout the projects?	Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.) Are those also following FAIR principles?	Aim to publish documents in Zenodo.
Allocation of resources		
Who will be responsible for data management in your WP/Task	ACTRIS DC	
How will long-term preservation be ensured?	By connection to ACTRIS bodies, where ACTRIS is a European Research Infrastructure	
Data security		
What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?	ACTRIS operated trusted data repositories, including backup and off-site backup.	
Will the data be safely stored in trusted repositories for long-term preservation and curation?	Yes.	
Ethical aspects		
Are there, or could there be, any ethics or legal issues that can have an impact on data sharing?	Yes or No. (If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action DoA).	No.
Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?	Yes or No.	Yes.

Other issues	
Do you, or will you, make use of other national/funder/sectorial/departm ental procedures for data management? If yes, which ones?	No.

AnaEE ERIC

RI Name	AnaEE ERIC	
Contact	Dario De Nart (dario.denart@crea.gov.it)	
Data Summary (Optional)		
Will you re-use any existing data and what will you re-use it for?	Metadata of AnaEE platforms recognised as interesting for EHN will be reused in the CoS	
Will you re-use any existing data and will this generate new data?	No	
What types and formats of data will the project generate or re-use?	Pre-existing ECVs data series	
What is the purpose of the data generation or re-use and its relation to the objectives of the project?	Demonstrate ENVRI-Hub integration and data access capabilities	
What is the expected size of the data that you intend to generate or re-use?	A few GBs	
What is the origin/provenance of the data, either generated or re-used?	Projects hosted on our platforms	
To whom might your data be useful ('data utility'), outside your project?	All environmental scientists and decision-makers	
FAIR Data		
1.) Making data findable, including provisions for metadata	<i>Will data be identified by a persistent identifier?</i>	URLs
	<i>Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.</i>	OpenAPI 3.0 descriptions
	<i>Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?</i>	Yes, you can find them in the OpenAPI specification

	<i>Will metadata be offered in such a way that it can be harvested and indexed?</i>	Yes, it's OpenAPI 3.0
2.) Making data openly accessible		
a) Repository:	<i>Will the data be deposited in a trusted repository?</i>	No, it lives on our servers
	<i>Have you explored appropriate arrangements with the identified repository where your data will be deposited?</i>	We pay for its maintenance with membership fees, so as long as Italy, France, Denmark, Belgium, and the Czech Republic don't pull the plug, everything should be fine
	<i>Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?</i>	Everything as a URL
b) Data:	<i>Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.</i>	Yes, you can already access it here https://developer.anaee.eu/
	<i>If an embargo is applied to give time to publish or seek the protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	No embargos
	<i>Will the data be accessible through a free and standardised access protocol?</i>	Yes, it's REST
	<i>If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?</i>	No, you can do whatever you fancy with it
	<i>How will the identity of the person accessing the data be ascertained?</i>	
	<i>Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?</i>	Nope

c) Metadata:	<i>Will metadata be made openly available and licensed under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?</i>	Yes, it's already there
	<i>How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?</i>	As long as AnaEE exists
	<i>Will documentation or reference about any software be needed to access or read or process the data be included? Will it be possible to include the relevant software (e.g. in open-source code)?</i>	There it is AnaEE services demo notebooks - Repos
3.) Making data interoperable	<i>What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?</i>	We assume people can read the documentation
	<i>In case it is unavoidable that you use uncommon or generate project-specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining, or extending them?</i>	We assume people can read the documentation
	<i>Will your data include qualified references to other data (e.g. other data from your project, or datasets from previous research)?</i>	No, it's self-contained
4.) Increase data re-use	<i>How will you provide the documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?</i>	There it is AnaEE services demo notebooks - Repos
	<i>Will your data be made openly available in the public domain to permit the widest re-use possible? Will your data be licensed using standard re-use licenses, in line with the obligations set out in the Grant Agreement? Under which license?</i>	Yes, it already is It's licensed under CC-BY 4.0

	<i>Will the data produced in the project be useable by third parties, in particular after the end of the project?</i>	No new data will be produced
	<i>Will the provenance of the data be thoroughly documented using the appropriate standards?</i>	Nope
	<i>Describe all relevant data quality assurance processes.</i>	Not sure if relevant, anyway, we do gap filling and other statistical data cleansing
Other research outputs		
In addition to the management of data, are you also considering and planning for the management of other research outputs that may be generated or re-used throughout the projects?	<i>Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.) Are those also following FAIR principles?</i>	We manage the VRE-Lib component of the ENVRI Hub
Allocation of resources		
Who will be responsible for data management in your WP/Task	Dario De Nart	
How will long-term preservation be ensured?	Until 2027	
Data security		
What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?	Continuous backup, geo-redundant storage	
Will the data be safely stored in trusted repositories for long-term preservation and curation?	<i>No, it's in AnaEE's cloud, which has no certification right now, but it works.</i>	
Ethical aspects		
Are there, or could there be, any ethics or legal issues that can have an impact on data sharing?		No

Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?		Yes
Other issues		
Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?	No	

EPOS ERIC

RI Name	EPOS ERIC	
Contact	Alessandro Turco (alessandro.turco@epos-eric.eu)	
Data Summary (Optional)		
EPOS ERIC manages the EPOS Platform that collects web services from different Earth science domains. Services relevant also in the EHN project will be made available in the CoS, and therefore, the data referenced by those services will be accessible to project partners. EPOS ERIC does not own the data provided by the aforementioned services. The CoS collects user data to serve AAI purposes and treats it in compliance with GDPR and other applicable regulations following EPOS Data Policy guidelines.		
Will you re-use any existing data and what will you re-use it for?	Metadata of EPOS platform services recognised as interesting for EHN will be reused in the CoS.	
Will you re-use any existing data and will this generate new data?	EPOS ERIC works on metadata, not on data. Data is provided by the National Research Infrastructures and Organisations and only “brokered” by EPOS.	
What types and formats of data will the project generate or re-use?	Metadata follows the EPOS-DCAT-AP standard or possibly project-related extensions of it.	
What is the purpose of the data generation or re-use and its relation to the objectives of the project?	Metadata helps scientists find, access, interoperate and reuse data made available by data providers.	
What is the expected size of the data that you intend to generate or re-use?	Metadata size is incomparably smaller than the data it refers to. Data size cannot be estimated from our side.	
What is the origin/provenance of the data, either generated or re-used?	Data providers are responsible for filling in the metadata descriptions of their datasets/services	
To whom might your data be useful ('data utility'), outside your project?	All environmental scientists and decision-makers	
FAIR Data		
1.) Making data findable, including provisions for metadata	Will data be identified by a persistent identifier?	EPOS integrates more than 300 different services, providing access to data. 70% of services provide access to data identified by persistent identifiers.
	Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.	EPOS requires data providers to describe their services with the EPOS-DCAT-AP standard.

	Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?	Yes
	Will metadata be offered in such a way that it can be harvested and indexed?	Metadata is exposed in a proprietary JSON format, which can be harvested
2.) Making data openly accessible		
a) Repository:	Will the data be deposited in a trusted repository?	It is up to the internal policies of the data providers. Some of them do not deposit data in trusted repositories.
	Have you explored appropriate arrangements with the identified repository where your data will be deposited?	It is up to the internal policies of the data providers.
	Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?	Same as above
b) Data:	Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.	- As a matter of fact, around 90% of the data linked to EPOS services is open. However, the EPOS Data Policy allows for three different possibilities: Open Access: Data freely available/ accessible to User(s) either for download or for direct use within an EPOS Service; Restricted Access: Data that is available under the conditions set out by the Data Providers. Restrictions to specific types of user categories, if any, should be limited to specific datasets. Restrictions may also mean that fees could be charged. While metadata shall always be available at no charge. Fees, if any, shall not be higher than the actual cost of making the Data available; Embargoed Access: Data that is available only after a predefined limited time (embargo period that should, as a general rule, last 6 months and can extend up to 3 years if motivated) has passed since the collection/generation of the Data. Once the embargo period has passed, the Data may become either Open or Restricted.

	<i>If an embargo is applied to give time to publish or seek the protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	See above
	<i>Will the data be accessible through a free and standardised access protocol?</i>	Yes, all datasets can be accessed by standardised (or de facto standard, as in the case of FDSN) access protocols.
	<i>If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?</i>	Through Authentication
	<i>How will the identity of the person accessing the data be ascertained?</i>	Through Authentication
	<i>Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?</i>	No
c) Metadata:	<i>Will metadata be made openly available and licensed under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?</i>	Yes
	<i>How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?</i>	Up to the data providers, ideally at least one decade (it is not linked to this project).
	<i>Will documentation or reference about any software be needed to access or read or process the data be included? Will it be possible to include the relevant software (e.g. in open-source code)?</i>	Yes, documentation is available.
3.) Making data interoperable	<i>What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?</i>	EPOS-DCAT-AP

	<i>In case it is unavoidable that you use uncommon or generate project-specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining, or extending them?</i>	EPOS-DCAT-AP is an already accepted, recognised and openly available extension of the DCAT-AP standard.
	<i>Will your data include qualified references to other data (e.g. other data from your project, or datasets from previous research)?</i>	Not in all cases.
4.) Increase data re-use	<i>How will you provide the documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?</i>	Depends on the datasets. Documentation on data is generally available, either as a PDF or as a weblink.
	<i>Will your data be made openly available in the public domain to permit the widest re-use possible? Will your data be licensed using standard re-use licenses, in line with the obligations set out in the Grant Agreement? Under which license?</i>	Yes, usually CC-BY-4 (depends on data providers)
	<i>Will the data produced in the project be useable by third parties, in particular after the end of the project?</i>	Yes
	<i>Will the provenance of the data be thoroughly documented using the appropriate standards?</i>	Depends on the dataset/service
	<i>Describe all relevant data quality assurance processes.</i>	Too many for this table, we have more than 300. Example: https://www.emidius.eu/AHEAD/data_quality_assurance.php from the "Bibliography for historical earthquakes" service
Other research outputs		
In addition to the management of data, are you also considering and planning for the management of other research outputs that may be generated or re-used throughout the projects?	<i>Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.) Are those also following FAIR principles?</i>	We provide the software for the CoS. The source code is open and licensed under the GPL v3 license.
Allocation of resources		

Who will be responsible for data management in your WP/Task	EPOS ERIC is responsible for the personal data stored/processed by its software. Our DPO is Lucio Badiali. The scientific data/metadata accessible through the CoS are under the responsibility of the data/service providers.	
How will long-term preservation be ensured?	Up to the communities/data providers.	
Data security		
What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?	Up to the communities/data providers.	
Will the data be safely stored in trusted repositories for long-term preservation and curation?	Up to the communities/data providers.	
Ethical aspects		
Are there, or could there be, any ethics or legal issues that can have an impact on data sharing?	<i>Yes or No. (If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action DoA).</i>	No.
Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?	<i>Yes or No.</i>	Yes.
Other issues		
Do you, or will you, make use of other national/funder/sectorial/depart mental procedures for data management? If yes, which ones?	No.	

Euro-Argo ERIC

RI Name	<i>Euro-Argo ERIC</i>
Contact	Delphine Dobler (Delphine.Dobler@euro-argo.eu)
Data Summary (Optional)	
Will you re-use any existing data and what will you re-use it for?	Yes, observational data from the Argo program and the associated access services will be further exposed on the federating platform ENVRI-Hub. Access means and outputs will be embedded within an ENVRI-applicable programming interface that manages access to all RIs' observational data in a homogenised way.
Will you re-use any existing data and will this generate new data?	Not per se (no computation from existing data, rather superimposition), but "data" can encompass other meanings. As data, services and documentation will be indexed for search in the knowledge base and services exposed on the catalogue of services, this will generate the necessary "metadata" for these ENVRI-HUB services.
What types and formats of data will the project generate or re-use?	Observational oceanic data and their associated metadata (spatio-temporal coordinates, sensor model, serial numbers, etc.)
What is the purpose of the data generation or re-use and its relation to the objectives of the project?	See above: expose Euro-Argo ERIC datasets and services in a federative web platform.
What is the expected size of the data that you intend to generate or re-use?	The data volume represents 2 TB of data and 4 million files. It is growing by approximately 600 files each day.
What is the origin/provenance of the data, either generated or re-used?	Data are generated within the international Argo program. Euro-Argo ERIC is responsible for the European contribution of this program (from the instrument deployment from ships to collection, processing and 3 different levels of quality controls (real-time, delayed-mode and basin-scale consistency). For each instrument (called a profiler float or float), the metadata indicates the entity responsible for the instrument itself and for the data generation.
To whom might your data be useful ('data utility'), outside your project?	Access to high-quality Argo data is essential for ocean and climate scientific studies, such as assessing the global characteristics and local variability of deoxygenation, ocean acidification, a better understanding of the carbon cycle and the anthropogenic carbon uptake by the ocean. It is noteworthy that most biogeochemical parameters are intertwined; let them be nitrate, Chlorophyll-a, oxygen, silicate, phosphate, pH, pCO₂, and light: all have their role in the carbon cycle. The Argo array is unique; it allows measurements to be performed all over the world's oceans, at any time of the year, when boats and very precise measurements are not possible. This allows understanding the characteristics and seasonal variability almost everywhere, even in regions and times where usual measures made from research vessels are not possible, e.g., the Polar regions. This autonomy of the Argo floats often comes with an inability to directly post-calibrate the sensor. This is the reason why continuous effort is made in defining and developing calibration procedures, using external but close in time and space high-quality data (as compared to the region variability and correlation scales), that are less numerous but more accurate.

The Argo dataset has been highly referenced in scientific publications since the inception of the Argo array (early 2000s). The number of peer-reviewed publications with a European main author is about 80 to 100 per year, on average, since 2014. At the international level, the number of publications per year is 300 to 400 (see https://www.euro-argo.eu/KPIs). Argo data are also thoroughly referenced in the 6th Intergovernmental Panel on Climate Change (IPCC) report (44 mentions, 320 publications and 4 key figures), and its importance for climate monitoring is highlighted in an infographic that can be viewed at https://infogram.com/argos-contribution-to-ipcc-reports-1h9i6q7pki5954g?live. In-situ Argo data are also integrated by operational users regarding ocean and ocean-atmosphere coupled models (European Copernicus Marine service reanalyses and other international services such as those from the USA/NOAA). Core parameters (temperature and salinity) are assimilated and significantly improve the accuracy of the models, this accuracy improvement being assessed through OSSE experiments, i.e. comparison of the model accuracy between various assimilation scenarios where all available data is assimilated, half of them, none of them. Operational users are eager to assimilate more widely BGC data to improve the reliability of environmental models. This step should be taken for the European Digital Twin of the Ocean (EDITO: https://www.edito.eu/). The in-situ accuracy and its accessibility play a crucial role in improving the accuracy of the Digital Twin of the Ocean reanalyses and predictive scenarios. In particular, the accessibility of high-quality BGC data improves the ability to derive as accurate as possible predictive scenarios.		
FAIR Data		
1.) Making data findable, including provisions for metadata	<i>Will data be identified by a persistent identifier?</i>	Argo data are persistently identified using a DOI: https://doi.org/10.17882/42182 . It includes monthly archives, which enables the citation of the exact dataset that was used in scientific studies and, thus, their reproducibility.
	<i>Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.</i>	Argo metadata vocabularies are web-exposed on the NVS. Any addition or change is internationally endorsed at the Argo Data Management level, following the CF convention.
	<i>Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?</i>	The discovery of the Argo dataset is enabled through various services. These services offer filtering options based on the Argo metadata. The two main graphical web services are the FleetMonitoring (https://fleetmonitoring.euro-argo.eu/) and the Argo Information Centre under the OceanOPS governance (https://www.ocean-ops.org/board?t=argo). Two other graphical web services allow data download: the Argo data selection (https://dataselection.euro-argo.eu/) and the ERDDAP service (https://erddap.ifremer.fr/erddap/tabledap/ArgoFlots.html).

		https://erddap.ifremer.fr/erddap/tabledap/ArgoFloats-synthetic-BGC.html).
	<i>Will metadata be offered in such a way that it can be harvested and indexed?</i>	Several harvesting services are possible: • The synchronisation service (http://www.argodatamgt.org/Access-to-data/Argo-GDAC-synchronization-service) • The indexation of files, as mentioned above, can be accessed through several technologies, from the simplest .csv (tabular) file to the web queries on a SparQL endpoint and the more recent STAC. The use of these various technologies is complementary as they target different kinds of users.
2.) Making data openly accessible		
a) Repository:	<i>Will the data be deposited in a trusted repository?</i>	Argo data are deposited in a trusted repository, which encompasses hot and cold redundancy mechanisms. Preservation is managed by Ifremer. Consistent with Argo data management policy, the US NODC ensures the long-term preservation of the Argo data.
	<i>Have you explored appropriate arrangements with the identified repository where your data will be deposited?</i>	N/A
	<i>Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?</i>	As mentioned above, a persistent identifier is associated with the dataset.
b) Data:	<i>Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.</i>	The Argo data and metadata policy is open and free, with no retention, and accessible. They are licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/), which means they can be freely shared and adapted, provided that the appropriate credit is mentioned and any adaptation is indicated. Data and metadata travel together. As mentioned above, there is a long-term preservation scheme including redundancy and sustained funding for that. Documentation is persistently identified, and the Argo user manual includes the necessary information to read and use the data (http://dx.doi.org/10.13155/29825). The other documents (http://www.argodatamgt.org/Documentation) inform about the processes used in the data workflow. The numerous software used along the data value chain are openly accessible. This is mainly performed on the GitHub web platform and its open version-control and management service. Their location on the web is described in the appropriate documentation. Euro-Argo ERIC

		manages its own GitHub organisation (https://github.com/euroargodev/). Finally, a user-friendly Python library (https://github.com/euroargodev/argopy) has been developed to access, subset and display Argo data from any operating system, among other features.
	<i>If an embargo is applied to give time to publish or seek the protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	No embargo
	<i>Will the data be accessible through a free and standardised access protocol?</i>	Yes (see above)
	<i>If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?</i>	No restrictions
	<i>How will the identity of the person accessing the data be ascertained?</i>	Not applicable to Argo data
	<i>Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?</i>	No access to personal/sensitive data for this dataset of type observations of oceanic data.
c) Metadata:	<i>Will metadata be made openly available and licensed under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?</i>	Same as for data: CC BY 4.0
	<i>How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?</i>	The long-term accessibility (to both data and metadata) is sustained by the international and operational Argo program.
	<i>Will documentation or reference about any software be needed to access or read or process the data be included? Will it be possible to include the relevant software (e.g. in open-source code)?</i>	See the answer to the first question of the data section.

3.) Making data interoperable	<i>What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?</i>	The interoperability is ensured through vocabulary, file format, data access and indexing capabilities. Some additional details: • A web-exposed and constrained vocabulary (https://vocab.nerc.ac.uk/search_nvs/cvll?searchstr=argo&options=governance), with the parameters list being mapped to the more generic SeaDataNet equivalent table, allowing for I-adopt framework application and cross-domain mappings. The Argo Vocabulary Task team defines and updates this vocabulary under the umbrella of the international Argo Data Management Team governance (ADMT). When relevant, updates must be validated by the ADMT. The content and governance of Argo tables in the NVS are continuously enhanced; • A well-shared NetCDF file format and CF standards application; • A wide variety of data accesses, targeting different kinds of users (modellers, scientific researchers, web IT needs and for the research infrastructure's use), making it more interoperable for these various needs • A variety of indices. As a reminder, indexes allow for a subset of relevant observations from a set of criteria, for instance, select vertical profiles within a spatio-temporal boundary box. One of these indexes is a SPARQL end-point: the metadata is exported in a triple point database-like storage on the web, and SPARQL is a web database query language, like SQL, allowing access to the unitary file URLs matching the request criteria. This is published with the Argo ontology (https://co.ifremer.fr/co/argo-linked-data/doc/argo-floats.ttl). Additionally, there are the legacy ASCII .csv indexes and the newly implemented STAC (Spatio-Temporal Asset Catalog) index.
	<i>In case it is unavoidable that you use uncommon or generate project-specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining, or extending them?</i>	One of the aims of the ENVRI-HUB's next project is the interface with common vocabularies, and this will be performed in the framework of the NVS.
	<i>Will your data include qualified references to other data (e.g. other data from your project, or datasets from previous research)?</i>	Not really applicable

4.) Increase data re-use	<i>How will you provide the documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?</i>	A lot of material is already provided above for the data re-use (license, documentation). The whole data processing chain is well described in official and openly published documents.
	<i>Will your data be made openly available in the public domain to permit the widest re-use possible? Will your data be licensed using standard re-use licenses, in line with the obligations set out in the Grant Agreement? Under which license?</i>	Yes
	<i>Will the data produced in the project be usable by third parties, in particular after the end of the project?</i>	Yes
	<i>Will the provenance of the data be thoroughly documented using the appropriate standards?</i>	Yes
	<i>Describe all relevant data quality assurance processes.</i>	One aspect of quality assurance is performed at the level of the measure/data (quality control methods) and at the level of metadata (controlled vocabularies). Additionally, the format and content of Argo NetCDF files are checked upon submission to the GDAC (FileChecker: https://github.com/OneArgo/ArgoFormatChecker). Regarding the management of activities (either issues or evolution needs), externally accessible follow-up tools are currently used: GitHub ticket system (both for the vocabularies and for other data format actions) and GitLab (for the IT structure, including GDAC and services). Additionally, the Argo data management team meets every year and, at this occasion, discusses/adapts/votes procedures and format where relevant.
Other research outputs		
In addition to the management of data, are you also considering and planning for the management of other research outputs that may be generated or re-used throughout the projects?	<i>Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.) Are those also following FAIR principles?</i>	Within the Euro-Argo ERIC, the software produced is published following the existing rules of the community, i.e. using the GitHub public repositories: https://github.com/OneArgo https://github.com/euroargodev https://github.com/ArgoDMQC They will be persistently identified using Zenodo and archived with Software Heritage within the current Euro-Argo One project (some software is already aligned, but a thorough review is foreseen in this frame).

		As already mentioned above, the documentation for data workflows and governance is exposed following the existing rules of the community, i.e. on the Argo Data Management Team website (http://www.argodatamgt.org/Documentation/) and the Euro-Argo website (www.euro-argo.eu), and persistently identified using DOIs. This material will also be harvested to be visible by searching in the ENVRI-HUB knowledge base.
Allocation of resources		
Who will be responsible for data management in your WP/Task	All costs for the allocation of resources are covered by the recurrent fund of the responsible entity.	
How will long-term preservation be ensured?	The long-term preservation will follow what is already in place.	
Data security		
What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?	The Argo dataset preservation will follow what is already in place, including warm and cold restart mechanisms for the processing chain as well as two long-term preservation schemas (monthly archives on SEANOE and long-term preservation by US NODC).	
Will the data be safely stored in trusted repositories for long-term preservation and curation?	Additionally, a cybersecurity analysis will be performed in the frame of the Euro-Argo ONE project to identify remaining vulnerabilities and take the necessary actions	
Ethical aspects		
Are there, or could there be, any ethics or legal issues that can have an impact on data sharing?	Yes or No. (If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action DoA).	Not applicable to the Euro-Argo data section.
Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?	Yes or No.	Same as above
Other issues		
Do you, or will you, make use of other national/funder/sectorial/departamental procedures for data management? If yes, which ones?	Not applicable	

RI Name	 	
Contact	Ulrich Bundke (u.bundke@fz-juelich.de) Andreas Petzold (a.petzold@fz-juelich.de)	
Data Summary (Optional)		
Will you re-use any existing data and what will you re-use it for?	Yes, metadata and data access services from IAGOS will be accessible through the CoS.	
Will you re-use any existing data and will this generate new data?	No new data will be generated.	
What types and formats of data will the project generate or re-use?	IAGOS observational data and associated averaged products in the NetCDF format	
What is the purpose of the data generation or re-use and its relation to the objectives of the project?	Improve IAGOS data discovery	
What is the expected size of the data that you intend to generate or re-use?	< 1 To Stored at the IAGOS Data Centre	
What is the origin/provenance of the data, either generated or re-used?	Data generated by IAGOS RI	
To whom might your data be useful ('data utility'), outside your project?	Environmental researchers, modellers, policymakers, etc.	
FAIR Data		
1.) Making data findable, including provisions for metadata	<i>Will data be identified by a persistent identifier?</i>	yes
	<i>Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.</i>	Yes Metadata existing at the IAGOS data centre Rich standard content: description, provenance, etc. ISO 19139
	<i>Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?</i>	Yes

	<i>Will metadata be offered in such a way that it can be harvested and indexed?</i>	Yes CSW
2.) Making data openly accessible		
a) Repository:	<i>Will the data be deposited in a trusted repository?</i>	Stored at IAGOS Data Centre (CoreTrustSeal certification in progress)
	<i>Have you explored appropriate arrangements with the identified repository where your data will be deposited?</i>	N/A
	<i>Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?</i>	Yes
b) Data:	<i>Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.</i>	Averaged products are fully open Observational datasets require authentication. Everyone can ask for access and will get it. The only reason is to be able to compile rich statistics of usage of the data. All data are available under a CC BY 4.0 license.
	<i>If an embargo is applied to give time to publish or seek the protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	N/A
	<i>Will the data be accessible through a free and standardised access protocol?</i>	https
	<i>If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?</i>	N/A
	<i>How will the identity of the person accessing the data be ascertained?</i>	Email
	<i>Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?</i>	No

c) Metadata:	<i>Will metadata be made openly available and licensed under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?</i>	Yes.
	<i>How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?</i>	As long as IAGOS and AERIS exist. Yes.
	<i>Will documentation or reference about any software be needed to access or read or process the data be included? Will it be possible to include the relevant software (e.g. in open-source code)?</i>	Data files are self-described
3.) Making data interoperable	<i>What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?</i>	Data follows CF and ACDD conventions for NetCDF Metadata in ISO 19115.
	<i>In case it is unavoidable that you use uncommon or generate project-specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining, or extending them?</i>	Yes. Usage of AERIS vocabulary mapped to standard vocabularies (CF standard names, GCMD, ECV).
	<i>Will your data include qualified references to other data (e.g. other data from your project, or datasets from previous research)?</i>	References to past versions.
4.) Increase data re-use	<i>How will you provide the documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?</i>	Documentation is available on the IAGOS Data Portal.
	<i>Will your data be made openly available in the public domain to permit the widest re-use possible? Will your data be licensed using standard re-use licenses, in line with the obligations set out in the Grant Agreement? Under which license?</i>	Yes CC-BY-4.0

	Will the data produced in the project be useable by third parties, in particular after the end of the project?	Yes
	Will the provenance of the data be thoroughly documented using the appropriate standards?	Yes
	Describe all relevant data quality assurance processes.	In the IAGOS DMP: https://permalink.aeris-data.fr/iagos_dmp_v1
Other research outputs		
In addition to the management of data, are you also considering and planning for the management of other research outputs that may be generated or re-used throughout the projects?	Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.) Are those also following FAIR principles?	Yes, software Following FAIR principles.
Allocation of resources		
Who will be responsible for data management in your WP/Task	IAGOS Data Centre; AERIS-CNRS France	
How will long-term preservation be ensured?	By IAGOS Data Centre (and AERIS-CNRS).	
Data security		
What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?	Archiving by AERIS-CNRS.	
Will the data be safely stored in trusted repositories for long-term preservation and curation?	Yes (CoreTrustSeal certification of IAGOS Data Centre in progress).	
Ethical aspects		
Are there, or could there be, any ethics or legal issues that can have an impact on data sharing?	Yes or No. (If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action DoA).	No

Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?	Yes or No.	No (N/A)
Other issues		
Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?	No	

ICOS ERIC

RI Name	ICOS ERIC	
Contact	Alex Vermeulen (alex.vermeulen@nateko.lu.se)	
Data Summary (Optional)		
Will you re-use any existing data and what will you re-use it for?	Metadata and data access services from ICOS will be accessible through the CoS. Metadata about relevant ICOS datasets will be mapped to DCAT and made available through the FDP.	
Will you re-use any existing data and will this generate new data?	No new data will be generated. All data access will ultimately happen through ICOS services.	
What types and formats of data will the project generate or re-use?	Observational atmospheric, ecosystem and oceanic data related to greenhouse gases. Results from various models which are related to the three observational themes. Most commonly used formats are NetCDF and CSV.	
What is the purpose of the data generation or re-use and its relation to the objectives of the project?	A major objective of the infrastructure is to provide unified access to metadata and data from various ENVRI. Providing access to ICOS services through the CoS directly serves this objective. So does providing access to metadata through the FDP.	
What is the expected size of the data that you intend to generate or re-use?	No data will be stored directly in EHN-related infrastructures.	
What is the origin/provenance of the data, either generated or re-used?	The core ICOS data are generated at ICOS-labelled stations and are the most relevant for EHN. Other data hosted at the ICOS Carbon Portal have various origins.	
To whom might your data be useful ('data utility'), outside your project?	ICOS data cover atmospheric, ecosystem and oceanic observations related to greenhouse gases, as well as results of models related to these three themes. Such data can be of interest to researchers, modellers, policymakers, etc.	
FAIR Data		
1.) Making data findable, including provisions for metadata	<i>Will data be identified by a persistent identifier?</i>	Yes, which is based on the SHA-256 hashsum of the data.
	<i>Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.</i>	Metadata follows an internal model which includes the list of authors, a description of the data, the variables contained in the data file, information about data acquisition (including start and stop time, and spatial coverage) and production of the data file, links to previous and future versions of the data, and more.
	<i>Will search keywords be provided in the metadata to optimize the</i>	Yes

	<i>possibility for discovery and then potential re-use?</i>	
	<i>Will metadata be offered in such a way that it can be harvested and indexed?</i>	Metadata can be harvested via a SPARQL endpoint. Metadata about individual datasets is available in JSON, Schema.org , RDF/XML, RDF/Turtle and ISO 19115 formats.
2.) Making data openly accessible		
a) Repository:	<i>Will the data be deposited in a trusted repository?</i>	Yes, the ICOS Carbon Portal.
	<i>Have you explored appropriate arrangements with the identified repository where your data will be deposited?</i>	Not relevant as we are the repository.
	<i>Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?</i>	Yes.
b) Data:	<i>Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.</i>	All data is available under a CC BY 4.0 license.
	<i>If an embargo is applied to give time to publish or seek the protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	Embargoes are possible by exception only, and time period is to be specified by the data provider, following the rules that apply to their specific dataset
	<i>Will the data be accessible through a free and standardised access protocol?</i>	Yes, direct HTTP access.
	<i>If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?</i>	No restrictions.
	<i>How will the identity of the person accessing the data be ascertained?</i>	Only the geo-location derived from their IP address.

	<i>Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?</i>	No.
c) Metadata:	<i>Will metadata be made openly available and licensed under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?</i>	All metadata is openly available.
	<i>How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?</i>	As long as ICOS exists and possibly longer.
	<i>Will documentation or reference about any software be needed to access or read or process the data be included? Will it be possible to include the relevant software (e.g. in open-source code)?</i>	Metadata and data are accessible and understandable by themselves, without the need for additional software. Documentation about specialized tools developed at the ICOS Carbon Portal is available in our GitHub repositories.
3.) Making data interoperable	<i>What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?</i>	A subset of the metadata is available in the ISO 19115 format. The full metadata follows our own vocabulary, but it will be mapped to the DCAT v3 and iAdopt vocabulary in the FAIR Data Point.
	<i>In case it is unavoidable that you use uncommon or generate project-specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining, or extending them?</i>	Our ontology is openly published. In addition, as mentioned above, it will be mapped to the DCAT vocabulary in the FAIR Data Point.
	<i>Will your data include qualified references to other data (e.g. other data from your project, or datasets from previous research)?</i>	Previous and next versions of datasets are part of the metadata.
4.) Increase data re-use	<i>How will you provide the documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?</i>	Some of these aspects are covered by scientific publications. Others can be provided by the ICOS thematic centers on request.

	Will your data be made openly available in the public domain to permit the widest re-use possible? Will your data be licensed using standard re-use licenses, in line with the obligations set out in the Grant Agreement? Under which license?	Yes, CC BY 4.0
	Will the data produced in the project be useable by third parties, in particular after the end of the project?	Yes.
	Will the provenance of the data be thoroughly documented using the appropriate standards?	Yes.
	Describe all relevant data quality assurance processes.	https://www.icos-cp.eu/data-services/data-collection/data-flow
Other research outputs		
In addition to the management of data, are you also considering and planning for the management of other research outputs that may be generated or re-used throughout the projects?	Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.) Are those also following FAIR principles?	ICOS is responsible for setting up the EHN FAIR Data Point, the software of which has been developed externally.
Allocation of resources		
Who will be responsible for data management in your WP/Task	Carbon Portal employees	
How will long-term preservation be ensured?	Metadata and data are stored at the ICOS Carbon Portal and are therefore expected to be preserved in the long term. The metadata in the FDP will be maintained by each individual research infrastructure and their long-term preservation depends on each of them. The FDP itself could be hosted by ICOS if needed in the long term.	
Data security		
What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?	Data is stored on our own secure servers. An external trusted repository used for all digital objects is B2SAFE; our instance is hosted by IT4I in the Czech Republic. Recovery is possible from B2SAFE, daily backups of our metadata and separate backups of our data.	
Will the data be safely stored in trusted repositories for long-term preservation and curation?	Yes, the ICOS Carbon Portal repository and B2SAFE.	
Ethical aspects		

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing?	<i>Yes or No. (If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action DoA).</i>	No
Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?	<i>Yes or No.</i>	No, not relevant
Other issues		
Do you, or will you, make use of other national/funder/sectorial/departamental procedures for data management? If yes, which ones?	No	

LifeWatch ERIC

RI Name	LifeWatch ERIC	
Contact	Lucia Vaira (lucia.vaira@lifewatch.eu) Eleonora Parisi (eleonora.parisi@lifewatch.eu)	
Data Summary (Optional) LifeWatch ERIC manages two main catalogues/repositories that collect all LifeWatch ERIC assets in the biodiversity and ecosystem domain, which are the LifeWatch ERIC Metadata Catalogue (metadatacatalogue.lifewatch.eu) and EcoPortal (ecoportal.lifewatch.eu) Metadata of LifeWatch ERIC assets that can be interesting for the EHN project will be reused in the CoS since it includes these two catalogues/repositories. LifeWatch ERIC does not own all the data provided by these catalogues since they include assets coming from its 8 National Distributed Centres and from other external Research Infrastructures		
Will you re-use any existing data and what will you re-use it for?	Metadata of LifeWatch ERIC services recognised as interesting for EHN will be reused in the CoS.	
Will you re-use any existing data and will this generate new data?	LifeWatch ERIC works on metadata, not on data. Data is provided by the National Distributed Centres and other external Organisations.	
What types and formats of data will the project generate or re-use?	Biodiversity and ecosystem-oriented data and their associated metadata.	
What is the purpose of the data generation or re-use and its relation to the objectives of the project?	Metadata helps researchers and scientists discover, access, and reuse resources made available by metadata providers.	
What is the expected size of the data that you intend to generate or re-use?	Metadata size is incomparably smaller than the data it refers to. Data size cannot be estimated from our side.	
What is the origin/provenance of the data, either generated or re-used?	Metadata providers are responsible for filling in the metadata descriptions of their resources.	
To whom might your data be useful ('data utility'), outside your project?	Researchers and scientists working in the biodiversity and ecosystem domain and decision-makers.	
FAIR Data		
1.) Making data findable, including provisions for metadata	<i>Will data be identified by a persistent identifier?</i>	URLs, URIs and DOIs.
	<i>Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.</i>	All LifeWatch ERIC resources are described by appropriate metadata schemas. In particular: ● Datasets: a subset schema of the EML 2.2.0 standard (77 fields); ● Services: a subset schema of the ISO 19139 standard (34 fields); ● VREs: a subset schema of the ISO 19139 standard (24 fields); ● Workflows: a subset schema of the ISO 19139 standard (24 fields);

		• Research sites: a subset schema of the ISO 19139 standard (24 fields); • Training resources: schema based on the EOSC training profile (22 fields); • Semantic resources: schema based on Metadata Vocabulary for Ontology Description and Publication (106 fields).
	<i>Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?</i>	Yes.
	<i>Will metadata be offered in such a way that it can be harvested and indexed?</i>	Yes, metadata is exposed in JSON and XML and can be automatically harvested.
2.) Making data openly accessible		
a) Repository:	<i>Will the data be deposited in a trusted repository?</i>	It is up to the internal policies of the data providers. Some of them do not deposit data in trusted repositories.
	<i>Have you explored appropriate arrangements with the identified repository where your data will be deposited?</i>	It is up to the internal policies of the data providers.
	<i>Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?</i>	URLs, URIs and DOIs.
b) Data:	<i>Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.</i>	Yes, according to the specific license.
	<i>If an embargo is applied to give time to publish or seek the protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	/

	<i>Will the data be accessible through a free and standardised access protocol?</i>	Yes: HTTPS for user interface access, REST APIs for machine access.
	<i>If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?</i>	Through authentication.
	<i>How will the identity of the person accessing the data be ascertained?</i>	According to the specific licence, authentication can be required or not to access data.
	<i>Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?</i>	No.
c) Metadata:	<i>Will metadata be made openly available and licensed under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?</i>	Yes.
	<i>How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?</i>	Yes, a metadata longevity plan is under definition, but in general, metadata will remain available even after the data is no longer available.
	<i>Will documentation or reference about any software be needed to access or read or process the data be included? Will it be possible to include the relevant software (e.g. in open-source code)?</i>	APIs are freely available for both the Metadata Catalogue (https://metadatacatalogue.lifewatch.eu/doc/api/index.html) and the EcoPortal (https://data.ecoportall.lifewatch.eu/)
3.) Making data interoperable	<i>What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?</i>	Metadata can be easily accessed via APIs to allow the reuse and exchange across RIs.
	<i>In case it is unavoidable that you use uncommon or generate project-specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining, or extending them?</i>	Yes, we could develop ad-hoc transformations when needed in order to have mappings.
	<i>Will your data include qualified references to other data (e.g. other data from your project, or datasets from previous research)?</i>	Not in general.

4.) Increase data re-use	<i>How will you provide the documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?</i>	Such information is self-included in metadata fields when possible.
	<i>Will your data be made openly available in the public domain to permit the widest re-use possible? Will your data be licensed using standard re-use licenses, in line with the obligations set out in the Grant Agreement? Under which license?</i>	Metadata records are freely available under a Public Domain License (CC0 1.0). The resources are usually licensed under CC-BY-4, but it depends on the data providers.
	<i>Will the data produced in the project be useable by third parties, in particular after the end of the project?</i>	Yes.
	<i>Will the provenance of the data be thoroughly documented using the appropriate standards?</i>	It depends on the dataset/service.
	<i>Describe all relevant data quality assurance processes.</i>	The publication of metadata records in both the Metadata Catalogue and the EcoPortal undergoes a deep review process that includes the involvement of expert reviewers.
Other research outputs		
In addition to the management of data, are you also considering and planning for the management of other research outputs that may be generated or re-used throughout the projects?	<i>Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.) Are those also following FAIR principles?</i>	We will continue to provide access to the ENVRI Training resources to have the ENVRI gateway up and running.
Allocation of resources		
Who will be responsible for data management in your WP/Task		
How will long-term preservation be ensured?	Up to the communities/data providers.	
Data security		
What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?	Up to the communities/data providers.	

Will the data be safely stored in trusted repositories for long-term preservation and curation?	Up to the communities/data providers.	
Ethical aspects		
Are there, or could there be, any ethics or legal issues that can have an impact on data sharing?	<i>Yes or No. (If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action DoA).</i>	No.
Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?	<i>Yes or No.</i>	Yes.
Other issues		
Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?	No.	

eLTER

RI Name	eLTER
Contact	Jan Bumberger (jan.bumberger@ufz.de) Martin Abbrent (martin.abbrent@ufz.de)
Data Summary (Optional)	
Will you re-use any existing data and what will you re-use it for?	eLTER will re-use existing legacy data sets that are and will be mobilized from the eLTER in-situ sites and eLTER platforms to allow access to data that are not published yet, as well as data that are already published at existing local, national, European and global platforms when it is related to the eLTER Sites and Platforms network and is related to the eLTER Standard Observations.
Will you re-use any existing data and will this generate new data?	Existing data will be reused to generate new data products with added value.
What types and formats of data will the project generate or re-use?	All types of data that are covered by the eLTER Standard Observations. This includes sensor-based timeseries data (tabular data), images, sound samples, in-situ samples, remote sensing data and high-throughput DNA sequencing data. It will generate model-based data ready for digital twin usage.
What is the purpose of the data generation or re-use and its relation to the objectives of the project?	eLTER aspires to provide Europe and the world with the scientific capacity to improve our understanding of terrestrial, freshwater, and transitional water ecosystems. One of the milestones to reach this aim is the provision of high-quality and reliable in-situ data and further added-value data integrated with modelling and remote sensing. In that context, one purpose is to enhance the data quality by harmonised data quality checks and enabling the creation of high-level FAIRified data products that combine data of different domains and spheres according to the eLTER Whole System Approach.
What is the expected size of the data that you intend to generate or re-use?	Depends on the context. The current 2024/25 eLTER Data Mobilisation Call produces approx. 70 GB of table-based data. Especially sound and camera-based sensor data, which are not yet operationally ingested into eLTER, will contribute significantly to the increase in data volume ingested in the future.
What is the origin/provenance of the data, either generated or re-used?	Data directly produced by the eLTER Sites and Platforms during campaigns and long-term observations. This data is managed and owned by the institutions that manage the eLTER Sites.
To whom might your data be useful ('data utility'), outside your project?	The data is useful for various actors. Among them are: 1. Peers in Research and Observation (sister Ris, Copernicus/ESA, WMO, GEO, UNECE, etc.); 2. Policy and decision-makers to allow for better decision-making (EEA, JRC,...); 3. Research and Science Stakeholders individual scientists, research-performing organisations, and research networks, to pertinent science initiatives at the international level; 4. Internal stakeholders like the Site and Platform Coordinators of eLTER.
FAIR Data	

1.) Making data findable, including provisions for metadata	<i>Will data be identified by a persistent identifier?</i>	Yes, DOI
	<i>Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.</i>	The eLTER Metadata Schemas are aligned to common schemas, including Dublin Core, EML, ISO19115, and Datacite. In addition, eLTER Metadata Schemas include elements to identify eLTER Standard Observations, the EUNIS Habitat and the eLTER Site (DEIMS-ID).
	<i>Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?</i>	Yes, it is possible to add individual keywords.
	<i>Will metadata be offered in such a way that it can be harvested and indexed?</i>	Yes, eLTER DAR provides REST-API and OGC CSW interface for harvesting metadata.
2.) Making data openly accessible		
a) Repository:	<i>Will the data be deposited in a trusted repository?</i>	eLTER RI will apply for CoreTrustSeal for the eLTER Digital Asset Registry (DAR) once eLTER ERIC is established.
	<i>Have you explored appropriate arrangements with the identified repository where your data will be deposited?</i>	eLTER DAR is developed and operated by the eLTER Topic Center for Data Management
	<i>Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?</i>	Datasets will be identified as digital objects by a DOI.
b) Data:	<i>Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.</i>	The eLTER data policy is not yet adopted, but eLTER follows the principle: "Data will be as open as possible, and as closed as necessary", following the FAIR principles. Restrictions and embargoes may apply and will be outlined in the data policy.

	<i>If an embargo is applied to give time to publish or seek the protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.</i>	This will be outlined in the data policy as soon as it is adopted. (max. 6 months).
	<i>Will the data be accessible through a free and standardised access protocol?</i>	Yes, e.g. OGC CSW.
	<i>If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?</i>	eLTER RI will provide long-term access to all data.
	<i>How will the identity of the person accessing the data be ascertained?</i>	Open access without identification. Restricted access controlled by AAI.
	<i>Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?</i>	A data provider or a team of data curators will be implemented to evaluate access to personal/sensitive data.
c) Metadata:	<i>Will metadata be made openly available and licensed under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?</i>	Metadata will be openly available and licensed under a public domain dedication.
	<i>How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?</i>	Forever.
	<i>Will documentation or reference about any software be needed to access or read or process the data be included? Will it be possible to include the relevant software (e.g. in open-source code)?</i>	It is generally possible to package software together with data. Common standard formats will be used to sustain accessibility, and FAIR Principles for research software will be applied and required.
3.) Making data interoperable	<i>What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?</i>	Depends on the data type, see above.

	<i>In case it is unavoidable that you use uncommon or generate project-specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining, or extending them?</i>	All vocabularies used in eLTER are published in the eLTER Vocab services. Controlled lists, vocabularies and ontologies will be registered in the common ontology and terminology services.
	<i>Will your data include qualified references to other data (e.g. other data from your project, or datasets from previous research)?</i>	Yes, the metadata schemas allow several options to cross-reference other datasets.
4.) Increase data re-use	<i>How will you provide the documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?</i>	Will be documented by the eLTER Standard Observation documentation that includes references, e.g. to data structure and format requirements. eLTER data products will also have clear documentation, including used codes (versions) to allow for re-usability.
	<i>Will your data be made openly available in the public domain to permit the widest re-use possible? Will your data be licensed using standard re-use licenses, in line with the obligations set out in the Grant Agreement? Under which license?</i>	In general, all data will be openly available under Creative Commons licences. In certain cases, there might be regulations and embargoes. This will be outlined in the data policy.
	<i>Will the data produced in the project be useable by third parties, in particular after the end of the project?</i>	eLTER RI will be a long-term research infrastructure with data management services designed for sustainability beyond project runtimes.
	<i>Will the provenance of the data be thoroughly documented using the appropriate standards?</i>	Yes.
	<i>Describe all relevant data quality assurance processes.</i>	Depends on the data type. ● Metadata Checks; ● Data format Checks; ● Data type-specific checks (from basic outlier checks to complex data analysis, including external data sources); ● Approval by data curators.
Other research outputs		

In addition to the management of data, are you also considering and planning for the management of other research outputs that may be generated or re-used throughout the projects?	<i>Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.) Are those also following FAIR principles?</i>	eLTER RI will mostly work with digital outputs. Requirements for storing physical assets will be covered by specific eLTER Topic Centers.
Allocation of resources		
Who will be responsible for data management in your WP/Task	Depends on the context: Jan Bumberger (jan.bumberger@ufz.de) Martin Abbrent (martin.abbrent@ufz.de)	
How will long-term preservation be ensured?	By sustaining the financing of the development, operation and maintenance of the eLTER RI and its repositories through the member countries of eLTER.	
Data security		
What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?	Data security is provided by the eLTER Topic Centre Data Management. Includes geo-redundant backups and a technical backbone with regular security assessments according to common standards.	
Will the data be safely stored in trusted repositories for long-term preservation and curation?	eLTER RI will apply for CoreTrustSeal for the eLTER Digital Asset Registry (DAR) once eLTER ERIC is established.	
Ethical aspects		
Are there, or could there be, any ethics or legal issues that can have an impact on data sharing?	<i>Yes or No. (If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action DoA).</i>	If there are issues, they will be outlined in the eLTER data policy.
Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?	<i>Yes or No.</i>	Yes.
Other issues		
Do you, or will you, make use of other national/funder/sectorial/departm ental procedures for data management? If yes, which ones?	Depends on the data type and sources. eLTER plans to include other local or national networks and research infrastructures by harvesting to avoid data duplication. In this case, this will include the specific procedures.	

SeaDataNet

RI Name	SeaDataNet
Contact	Peter Thijsse (peter@maris.nl)
Data Summary (Optional)	
Will you re-use any existing data and what will you re-use it for?	SeaDataNet is a data aggregator of validated marine observation datasets held by the main marine data centers in Europe. These datasets are originally collected by hundreds of originators, like research organisations and monitoring agencies, and are offered via SeaDataNet for re-use to science, industry, consultants, etc. The data is also reused within the SDN network and in EMODnet to generate data products like long-term climatologies.
Will you re-use any existing data and will this generate new data?	Data is validated and made more FAIR, leading to new versions of the data and metadata. Plus is used in data products. Within ENVRI-Hub NEXT, the datasets might be cached and used for processing/subsetting in combination with datasets from other RIs
What types and formats of data will the project generate or re-use?	ODV/ASCII/NetCDF for the datasets, and NetCDF for the products.
What is the purpose of the data generation or re-use and its relation to the objectives of the project?	See above.
What is the expected size of the data that you intend to generate or re-use?	The total of datasets in SDN is in the hundreds of GB, excluding seismic datasets that are much larger. The data products are not large either.
What is the origin/provenance of the data, either generated or re-used?	Research organisations, universities, monitoring agencies and industry. The origin is marked in the metadata.
To whom might your data be useful ('data utility'), outside your project?	Long-term climatologies, environmental conditions, trends, to researchers, consultants, governments and more.
Allocation of resources	
Who will be responsible for data management in your WP/Task	MARIS/IFREMER
How will long-term preservation be ensured?	Original versions of the datasets are always kept in the national data centres. And SeaDataNet itself also stores the versions of datasets and data products on its technical infrastructure and SEANOE with a DOI.
Data security	
What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?	Data: stored in national data centers, with replication versions in the SDN cloud. Data can always be replicated and loaded to the cloud. The central servers are professionally secured (CINECA/MARIS/IFREMER). The environmental data cached on the servers is CC BY licensed and not sensitive. Sensitive data is held by the datacenters, with other license types, and is only released upon user request (and contact).

Will the data be safely stored in trusted repositories for long-term preservation and curation?	Data: stored in national data centers, with replication versions in the SDN cloud. Products: SEANOE/ZENODO	
Ethical aspects		
Are there, or could there be, any ethics or legal issues that can have an impact on data sharing?	Yes or No. (If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action DoA).	No, a data policy and user license is provided. https://www.seadatanet.org/Data-Access/Data-policy
Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?	Yes or No.	Consent is requested before contacting/adding to mailing lists.
Other issues		
Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?	No	