ENVIRONMENT RESEARCH INFRASTRUCTURES
INNOVATION ROADMAP

ENVRINNOV

HORIZON-INFRA-2023-DEV-01

ENVRINNOV | GRANT NO. 101131426

D7.2- Refined Data Management Plan (DMP)

30/06/2024

Funded by the European Union

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

Deliverable Number	Deliverable Title	Lead Beneficiary	Type	Dissemination Level	Due Date (in months)
D7.2	Refined DMP	1 – CYI	Report	PU	6

Version	Date	Changed page(s)	Cause of change	Partner
V1	10/04/2024	Initial version	-	CYI
V2	18/06/2024	throughout	Input and comments from Consortium	All
V3	28/06/2024	throughout	Review and finalization by coordinator	CYI

Table of Contents

1. Definitions

Term	Definition
ENVRI	the network of European Research Infrastructures in the Environment Domain.
ENVRI-FAIR	ENVRI-FAIR is a recently completed H2020 funded project, aimed at advancing the findability, accessibility, interoperability, and reusability (FAIRness) of the data and services offered by the ENVRI community and to connect them to the emerging European Open Science Cloud.
ENVRI-HUB	the ENVRI-HUB has started being developed within the framework of the ENVRI-FAIR project and continues through the recently funded Horizon Europe project ENVRI-HUB NEXT to serve as a central gateway to environmental data and services offered by the European environmental research infrastructures. ENVRI-Hub will be part of EOSC. Data will be open and free to use by anyone.
EOSC	European Open Science Cloud, a European Commission initiative aiming at developing an infrastructure providing its users with services promoting open science practices.
Metadata	a set of data that describes and gives information about other data.
Creative Commons (CC) Licenses	a standardized way to grant the public permission to use creative work under copyright law.
Embargo Period	is a predetermined period during which access to a dataset or publication is not yet publicly available.

2. Introduction

This deliverable outlines the detailed Data Management Plan for the ENVRINNOV project. It builds on the initial overview of DMP contents as this was outlined at proposal stage (section 1.2.5 Open Science).

The DMP's aim is to ensure that the project manages data responsibly, in line with FAIR principles, and following the principle of 'as open as possible as closed as necessary', to identify and appropriately manage any necessary open access research data exceptions.

Accordingly, this DMP has been developed according to EU guidelines for Horizon Europe projects that produce, collect or process research data as these are outlined in the Horizon Europe DMP template. In line with these guidelines, the version of the DMP for ENVRINNOV presented in this document, is not meant to be final, nor does it present comprehensive answers to all the questions outlined in the Horizon Europe DMP template. Instead, the DMP is a living document, which will be reviewed and updated in regular intervals throughout the lifespan of the project. It is expected that more, and more granular information, will be integrated in the DMP, as the implementation of the project progresses and relevant activities within it evolve. In line with that, updates of the DMP are foreseen in the context of the Periodic Reports to be delivered to the European Commission for project review purposes.

The version of the DMP outlined in this document, and the processes and practices it outlines, have been influenced by the wider breadth of resources and tools available for best practices in data management for Horizon Europe projects, such as the Research Data Alliance's Metadata Standards Directory for its repository of discipline-specific standards and associated tools, and the Guide for Researchers on compliance with the Horizon Europe mandate for Research Data Management available in OpenAIRE.

Finally, this DMP also builds on existing results and best practices developed by the ENVRI community, primarily within the framework of the ENVRI-FAIR project, and its applications in ENVRI-HUB, including links to EOSC.

3. Data Summary

During the project, the following types of data are foreseen to be generated and/or re-used. The potential formats they will entail, the purpose these will serve in relation to the objectives of the project are outlined through links to the WP activities to which they correspond to:

Data types	**Formats**	**Purpose of data** (Project WP contribution)	**Linked to Objectives***
Surveys (questionnaires and results), audio and/or video recordings of interviews, trainings, and/ or meetings, reports and/or policy documents. This may include personal data, in cases that data collection is not anonymized. (See section 8 of this document for personal data.)	CSV, XLX, DOC, PDF, WAV, MP4, MOV, AVI	WP1,3,5,6	Obj: 1, 2, 3
Experimental data gathered during the innovation strategy pilots. Including in T2.1.: 1) Atmospheric measurement data acquired during test, calibration, validation of the UAV-GHG; 2) Information on Technology (description of TRL progress); and in T2.4.: Manual for using the processing and visualization tool for CRNS data	T2.1.: 1) CSV, TXT, NETCDF; 2) DOC, PDF, PPT, 3) MP4 T2.4.: 1) DOC, PDF	WP2	Obj: 2
Websites, mailing lists and/or event signup lists (which may include personal data of subscribers/attendees). (See section 8 of this document for personal data.)	HTML, CSV, XLX	WP4,5,7	Obj: 1, 2, 3
Promotional material (including digital brochures and multi-media e.g. video)	PDF, JPEG, PNG, MP4, MOV, AVI	WP7	Obj: 3
Project documents, deliverable reports, digital objects (e.g. visual brand)	DOC, PDF, JPEG, PNG, SVG, EPS	All WPs	Obj: 1, 2, 3

*As outlined in the proposal (section 1.1) ENVRINNOV is driven by three objectives:

1. Conduct a comprehensive analysis of ENVRI services and ecosystem technological needs and gaps and define methodologies and tools for their regular monitoring and updating.

2. Define, test, validate, digitalize, and promote the uptake of common strategies to enhance innovation across the ENVRI community network.

3. Establish synergies and complementarities with relevant ecosystems, and plan for the successful, realistic, and timely implementation of the ENVRI Innovation Roadmap.

Estimated expected size of data**:** The consortium anticipates collecting, producing, and storing a few tens of gigabytes of data.

Re-use of existing data and purpose of re-use: Results of past (completed) or parallel (ongoing) EC-funded research projects (e.g. H2020 ENVRI-FAIR GA No. 824068, and other projects under the INFRA-DEV-01-05 call) will be used for reference, and to ensure synergies and complementarities. All relevant results from such projects are also to be adhering to FAIR, open access and Open Science principles.

Data utility: Raw data, made available through the ways outlined in section 4, will primarily be used by the scientific community. Through Dissemination, Exploitation and Communication (DEC) activities the consortium will, as appropriate, translate data into outputs tailored to wider project stakeholders including those in RIs, industry, policymakers and the public.

Data origin/provenance:
The origin of the experimental data aligned with the innovation strategy pilots of WP2 mentioned above, includes, for T2.1 data related to the technology development (TRL increase from 4 to 7) of the UAV-GHG such as 1) information on hardware/software model/brand of drone components and atmospheric sensors and their configuration as an integrated system, 2) test, calibration, validation of the different components (UAV and GHG) in the lab (ICOS-ATC) and in the field (ambient conditions), 3) use of various computer-based models (e.g. Lagrangian model) to process environmental data (e.g. concentrations, GPS positioning, 3D wind, T/RH) to geolocate methane leaks and calculate methane fluxes. Whilst as part of T2.4, a processing and visualization tool for CRNS data is being developed. This will be published as open-source, ready-to-use software. In addition, a video will be created that shows the use of the most common CRNS models and introduces the use of the software.
For the purposes of the landscape analysis required to complete activities of WP1, WP3, and WP5, open access data derived from past ENVRI and ESFRI projects will be utilized within the framework of desk research. These are accessible through the ENVRI Community Zenodo in the form of public deliverable documents.

4. FAIR Data

As a general principle, research data produced, collected, or processed in the context of ENVRINNOV, follow the FAIR principles, meaning that data is Findable, Accessible, Interoperable and Re-usable. How the project ensures this principle, has been defined in line with what is outlined in the Horizon Europe DMP template, as required, and informed. by further input on FAIR data principles available on the FORCE11 discussion forum on FAIR data, and the Global Open FAIR Initiative.

In line with the framework set for the ENVRI community by the H2020 project ENVRI-FARI, ENVRINNOV also builds on the principles developed by the ENVRI-FAIR projects, whose aim has been to advance the findability, accessibility, interoperability, and reusability (FAIRness) of the data and services offered by the **ENVRI** Cluster research infrastructures and to connect them to the emerging European Open Science Cloud.

4.1. Making data findable, including provisions for metadata

ENVRINNOV will not deal with environmental data relevant to RIs (e.g. observations of the atmosphere / hydrosphere / biosphere / geosphere), which are managed by respective RIs, and collectively organized via the ENVRI-HUB (ENVRI-FAIR, ENVRI HUB-NEXT) for the specific formatting and indexing of environmental and their open access in the respective ENVRIs data centres.

Except for WP2, ENVRINNOV will not generate scientific data that will require the provision of metadata and make them findable.

Data generated in WP2 are related to technological developments and therefore may have to be protected for further exploitation (e.g. commercialization) purposes. In the case, this data (generated in WP2) will lead to open-access scientific publications (with open access of data), the project will ensure consistent use of RICH (meta)data with standard (unique/persistent) identification mechanisms registered or indexed in a searchable resource (e.g. DOIs), trusted repositories (e.g. Zenodo), EOSC, ENVRI-Hub, and widely used formats/naming conventions.

ENVRINNOV does not aim to generate new research data / environmental data. In the unlikely event that relevant environmental data would be generated (e.g. through the scientific exploitation of technologies developed in WP2), environmental data will be formatted following the respective (meta)data format already implemented within the respective environmental RI and stored within the respective data centre, with appropriate keywords provided in the metadata to optimize discovery and re-use, as well as being able to be harvested and indexed.

4.2. Making data accessible

ENVRINNOV generated research data/outputs will be open access by default, deposited in trusted repositories, and made available as soon as possible.

Such trusted repositories include primarily Zenodo, which will automatically register a Digital Object Identifier (DOI) for a record once it is published. The DOI is a globally unique persistent identifier which ensures that the record can be uniquely cited which is important for reproducibility and attribution of credit.

Zenodo is where the ENVRI community retains a collection of documents and data from past community projects and activities. As per the curation policy of the collection, data in the collection need to be relevant to the wider ENVRI community, rather than a single research organization, they need to be of acceptable quality (as determined by the curators), original and must have relevant metadata for connection to the research infrastructures involved. Typical examples are deliverables from projects involving environmental research infrastructures, particularly ones involved in the ENVRI community (envri.eu), of which ENVRINNOV is a core project member.

All necessary exceptions to the above will be duly justified, for example if certain datasets cannot be shared or need to be shared under restricted access conditions, such in the case of data that are to be utilized for further exploitation (including commercialization), in which case opening data goes against the legitimate interests of the consortium.

Justifications for such cases will clearly separate legal and contractual reasons from international restrictions. Further, if an embargo period must be applied to give time to publish or seek protection of intellectual property, this will be specified in the justification, while measures will be taken to ensure that data will be made available as soon as possible these exceptions no longer apply.

The above will be monitored within the framework of Task 7.3 IP and Knowledge Management and Protection, which looks after IP and Knowledge Management and Protection for the project, including in terms of analysis of IP and results created through the project, as well as relevant protection measure recommendations, to better support its path to long-term impact. Key Exploitable Results (KERs) will be identified and monitored under this task. It will also support the creation of the project Results Ownership List (ROL), which will be provided in the project's final periodic report. Interim results, and any necessary exceptions to ENVRINNOV's open data approach, will be reported within the framework of the project's periodic reporting and D7.3 IP Landscape analysis, protection measures recommendations (M24).

Access Guidelines and Governance

Decision-making regarding the above, will be facilitated through the governance instrument of the ENVRINNOV Steering Committee, whose members will also act as Data Access Committee members, to determine, evaluate and approve access status for sensitive data sets, and evaluate/approve access requests as these arise. Such discussions will take place as part of the agenda of the Steering Committee's termly meetings.

Overarching guidelines regarding access and availability of data from innovation activities involving ENVRIs beyond the project duration, will be developed within the framework of Task 3.1, including regarding special situations involving "sensitive" data, such as those co-created with private partners.

The ENVRI-Hub, ENVRI and ENVRINNOV websites will also be used to enhance accessibility of project outputs. However, private data on project coordination/management will be made available only to the consortium and the European Commission as needed.

Metadata

Access to operational (meta)data (e.g. from case pilots in WP2) will be made retrievable by identifier using the standardized communication protocol defined in ENVRI-FAIR and followed by the ENVRI community.

4.3. Making data interoperable

By default, the project will strive for continuous compatibility with existing standards, formats, and vocabulary conventions, and in line with new products, services and protocols as defined for the ENVRI Community in ENVRI-FAIR and will continue to be refined within the framework of the ENVRI-HUB Next project (for specifics, see the ENVRI-FAIR D1.6: Final DMP at zenodo.org).

Further, a large volume of data (e.g. in WP2) already follows established protocols that are interoperable, such as EBAS.

The Consortium will use widely available tools and software (e.g. MS Office) for reports, deliverables, and results pertaining to WP1, WP3-7.

4.4. Increase data re-use

In terms of the reusability of data/research outputs, open access to data will be ensured by default, as soon as possible, under open licenses (e.g. CC BY) to permit the widest re-use possible, in line with the obligations set out in the Grant Agreement. Necessary exceptions (e.g. for exploitation purposes) will be duly justified, as outlined in section 3.1.

In the case environmental data (generated in WP2) would lead to open-access scientific publications (with open access of data), proper meta-data and documentation needed to validate data analysis and facilitate data re-use will be made available, whilst provenance of

the data will be thoroughly documented using the appropriate standards (as per Environmental RIs Data Centres structure). Through this approach for open-access scientific publications, and applicable to other types of scientific exploitation, data produced in the project will be useable by third parties, after the end of the project.

5. Other research outputs

In addition to the management of data, the Consortium also recognizes the need to consider and plan for the management of other research outputs to be generated or re-used throughout ENVRINNOV, either digital (e.g. any protocols or workflows that may result from WP3, and/or an online platform to be developed as part of WP4) or physical that may emerge as part of the project. As a principle, the management of these research outputs will follow the principles of FAIR data management outlined in this DMP. Any information about any research output, tool, or instrument needed to validate the conclusions of a publication, will be made available via a trusted repository. Any exceptions or deviations pertaining to these will be duly justified, and sufficient details on how they will be managed and shared, or made available for re-use, in line with FAIR. These provisions will be considered by all WPs as part of the planning of their activities and results and will be reported in Project Period Reports and/or Deliverables as pertinent. Updates to the DMP to include these conditions will be made as necessary and reported as part of ENVRINNOV periodic reviews.

6. Allocation of resources

Costs for ensuring data is FAIR, including direct and indirect costs related to storage, archiving, re-use, security will be covered by the Consortium members own funds, as appropriate.

Data management responsibility for the project will be assumed by the Data Manager of the Climate and Atmosphere Research Center (CARE-C) of the Cyprus Institute (the Coordinating Beneficiary).

Long term preservation will be ensured by using the established databases and trusted repositories mentioned above (e.g. via upload to EBAS). Beyond the project duration, discussions are also underway for project results to be maintained on the ENVRI-HUB (as per the ENVRI-HUB Data Management Policy). One potentiality currently being explored is for ENVRI-HUB Next to become a Cluster Node in the EOSC ecosystem. Further developments on this will be reported in future iterations of this DMP.

7. Data security

Cyl will assume responsibility to ensure that project data are stored securely in the coordinators' server, and provisions are made to ensure that data security extends beyond the lifespan of the project. This includes:

- limited writing access.
- account control system.
- transmission via dedicated VPN infrastructure and stored in high availability databases.
- multiple compute and storage nodes.

Data curation will be done via manual and automated methods. Both raw and curated data will be included in daily and monthly off-site backups ensuring their integrity (as per Cyl infrastructure's disaster recovery plan). All measurements are backed up in a server in near real time. The server creates a backup every day.

Open access outputs will be stored in trusted repositories (such as Zenodo) for long-term preservation and curation.

Beyond the duration for the project, and in the case project results migrate to the ENVRI-HUB and/or the ENVRI community website, security will be ensured by the hosts of the respective web platforms (EGI for ENVRI HUB and ICOS-ERIC for the ENVRI Community website).

8. Ethics

The ENRINNOV project will comply with all requirements of The General Data Protection Regulation (GDPR) and the Data Protection Act (DPA) 2018. Additionally, the Coordinator will engage its appointed institutional Data Protection Officer (DPO) for advice on Data Protection matters. These measures will help ensure full compliance with data privacy rules, appropriate stewardship, curation, and preservation (as appropriate) of all data and research materials, while appropriately acquiring informed consent, as necessary.

In the case of collection of personal data for communications purposes for WP7 and/or for the landscape analysis and assessment to be conducted as part of WP1, WP3 and WP5 which may require the conduct of surveys, audio or video recordings of interviews which may contain personal data, the respective consortium partner leading the data collection activities shall ensure full compliance with relevant data protection legislation (both GDPR and national legislations according to the partner country through compliance also with the privacy policy of the respective Consortium partner organization), and ensure that research participants consent to, and are fully informed of the purposes and process through which data is collected and managed. Personal data collected for the project implementation purposes noted above, that will need to be required to be stored, will be done so in the project's shared folder, securely hosted on the coordinator's (CYI) organizational SharePoint drive. The security of the shared folder is safeguarded by the policies of the Cloud provider (Microsoft), it is hosted in the EU, whilst access requires log-in with credentials of each respective project partner. No further information or metadata will be stored for personal data, and they will only be accessible by consortium members, on a need to basis, only for the purposes of project implementation. The retention period for personal data will coincide with the lifetime duration of the project. In case that personal data need to be retained beyond project duration, the reason for this will be duly justified, and data contributors will be duly informed as necessary. Personal data will not be subject to re-use, nor will they be interoperable.

The Consortium will also comply with Article 14 (Ethics and Values) of the Project Grant Agreement, while Consortium members are bound by their institutional ethics structures (such as Ethics Committees), while in general, the DMP principles are designed with the aim to ensure integrity, quality, and transparency. This will apply to data collection practices across the project's Work Packages, whereas researchers and participants will be made fully aware of the purpose and the rationale of the data collection, their own contribution, and the way in which the findings will be processed. In addition, as project coordinator, CyI shall ensure the consortium guarantees the treatment of personal data generated during the project based on ethical standards and requirements deriving from European Legislation (i.e. right of privacy), the Horizon Europe programme (i.e. do-no- significant harm principle etc.), the GDPR, the ALLEA European Code of Conduct for Research Integrity, and EARMA-ERION. All the above will be coordinated as horizontal for the Project via the dedicated Task 8.2 Quality Assurance, Risk Management and Ethics. Lead: CyI, Partners: All (M1-M36). A Plan for Quality Assurance, Risks Management and Ethics will be formulated to consolidate, record, and monitor the above, with a dedicated deliverable on Ethics Compliance (D8.3, Lead: CyI, M12).