‘European Higher Education Institutions Network for Climate and Atmospheric Sciences’

HORIZON – WIDERA-2021-ACCESS-05

EDU4CLIMATE | GRANT NO. 101071247

D5.2 Data Management Plan

April 2023

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 REA. 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)
D5.2	Data Management Plan	1 – CyI	Report	Public	M6

Version	Date	Changed page(s)	Cause of change	Partner
V1	10/01/2023	Initial version		CyI
V2	28/03/2023	Updated	Reviewed included from Coordinator Data Manager and relevant consortium stakeholders. Updates implemented by Coordinator.	CyI
V3	28/04/2022	Final Version	Reviewed by Coordinator	CyI

Disclaimer: The information in this document is subject to change without notice. Company or product names mentioned in this document may be trademarks or registered trademarks of their respective companies.

Table of Contents

Contents

1. Introduction

This deliverable outlines the detailed Data Management Plan (DMP) for Edu4Climate. It builds on the initial overview of DMP contents as this was outlined at proposal stage (section 1.2.2 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.

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 Edu4Climate 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 as it is outlined here, is meant to be 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 review purposes of the Edu4Climate project.

The version of the DMP outlined in this document, and the processes and practices it outlines, have also been influenced by the wider breadth of resources and tools available online for best practice 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 How to comply with Horizon Europe mandate for Research Data Management available in OpenAIRE.

Finally, the present version of the DMP has benefitted from the consolidated know-how of Edu4Climate Consortium Partners and has built on pre-established practices and processes established at the respective Institutions to support open research, innovation, and education activities.

2. Data Summary

Types and formats of data that Edu4climate will generate or re-use, and its purpose in relation to the objectives of the project, include:

Data Types	**Formats**	**Purpose of data** (Project WP contribution)	**Linked to Objectives***
Reports and other publications and documents	DOCX, PDF, XLSX, PPTX	All WPs; including Deliverables	All (1-4)
Simulation data and data formats found and used within the climate and atmospheric modelling community	GRIB, NetCDF, and HDF	WP2	2
Experimental atmospheric data gathered within the joint research & innovation projects from in-situ observations and ground-based/satellite-based remote sensing	·in-situ: NASA-AMES 1001, ASCII and other excel-compatible delimited formats ·ground-based/satellite-based remote sensing: NetCDF	WP2, WP3	2, 3
Processed data in the form of data visualizations, graphs, or other reports	Various image and video file formats incl. JPEG, PNG, MP4	All WPs	All (1-4)

*As outlined in the proposal (section 1.1) the Edu4Climate strategy is driven by four objectives as per below:

1. *Education & Training*: Through the development of a comprehensive set of High-Quality Education and Training activities, Edu4ClimAte will strengthen the profile and foster capacity building of the

consortium, with a long-term view of preparing the network towards the establishment of a European Universities Alliance.

2. *Research*: Edu4ClimAte will strengthen the atmospheric observations and simulation capability and capacity of widening partners towards a better understanding of emissions, transport, transformation, sinks, feedback loops and trends of air pollutant and GHG in the EMME.
3. *Innovation*: The technical and innovative capacities of widening partners will be strengthened with the development of new environmental products/services of high added value in close partnership with businesses.
4. *Networking*: Local, regional, and international clusters led by the Edu4ClimAte consortium will be strengthened by aligning them with the ambitions of the European Research and Education Areas.

Re-use of existing data and purpose of re-use: Results of previous or parallel research projects (e.g. H2020 Teaming EMME-CARE GA No. 856612) will be used for reference, and to ensure synergies and complementarities. All relevant results from these projects also adhering to FAIR, open access and Open Science principles.

Estimated expected size of data: Generally, the consortium will collect, produce, and store large volumes of data, e.g. from climate model, if considering post-processed data and other relevant datasets to be used for analyses, can reach hundreds of terabytes.

Data origin/provenance: The origin of the data is derived from in situ observations and numerical model simulations. The former include ground based measurements from the CAO network, vertical profiles using drone technology and ground based mobile measurements from mobile platforms.

Data utility: Raw data and model data outputs will primarily be of use to the research community. Post-processing and research exploitation activities will, when appropriate, translate data inputs into practice-oriented information visualized and structured at different levels of complexity in line with the needs of project stakeholders including business, government, citizens and society.

3. FAIR data

As a general principle, research data produced, collected or processed in the context of Edu4Climate, follow the FAIR principles, meaning that data is Findable, Accessible, Interoperable and Re-usable. How Edu4Climate ensures this principle, has been defined in line with what is outlined in the Horizon Europe DMP template, as well as has been informed by further input on FAIR data principles available on the FORCE11 discussion forum on FAIR data, and content made available by the Global Open FAIR Initiative.

3.1 Making Data Findable

Edu4Climate will be making Data Findable, including Provisions for Metadata, in the following ways:

Widely: through the use of metadata, and/or standard identification mechanisms, such as persistent and unique identifiers, as indicated in the FAIR template. Such data will typically arise from WPs relating to areas of research, education, training and innovation, and dissemination, exploitation and communication – e.g. dust/air quality forecasting made available through the EMME-AMP (WP2) and near-real- time local Air Quality data publicly available (WP5) are other examples of data that will be widely findable.

Data produced for the purposes of the project and project-relevant experiments will follow widely accepted and accessible naming conventions and keywords that make them easily discoverable, to optimize the possibility for discovery and potential re-use. For example, data gathered and produced by and within the Cyprus Atmospheric Observatory (CAO) context (connected to 2. Research and 4.Networking Project Objectives) follow ACTRIS naming conversions, providing search keywords to optimize possibilities for re-use

and follow the EBAS[1] protocol for clear version numbering. Each data file downloaded from EBAS includes a header protocol, enforced by ACTRIS, that ensures enough metadata, tailored to each variable, are provided to the user.

Additionally, rich metadata data will be provided to allow discovery, using well-established disciplinary standards. The EUDAT B2SHARE service will be used for this purpose, where the metadata of the shared data are published though the OAI-PMH standard interface (http://www.openarchives.org/pmh/), and are harvested and indexed and become findable through the EUDAT-B2FIND research data network. For example, observational data will be shared using the ICOS or elTER repositories and will be readily available through EUDAT-B2FIND service. Also, climate model outputs, produced during simulation experiments, include useful metadata information that follows the "CF (Climate and Forecast) metadata" NetCDF (Network Common Data Form) convention and standards (e.g., simulation experiment details are stored in the NetCDF output; https://cfconventions.org/) that helps make the data identifiable. If the model output is post-processed then additional metadata attributes (e.g., history of commands/scripts used, statistics applied etc.) will be appended to the modified dataset, in accord with the CF conventions so that the end-user will have a detailed information about the data product (similar to what is applied to the CMIP5/CMIP6 datasets).

On a restricted basis: for access only by the consortium partners and the European Commission services. These concern data that need to be treated as sensitive or confidential, primarily arising from WPs that relate to coordination and management, and infrastructure – e.g. the Initial and Final Action Plan (Tasks 4.1 and 4.4); as well as some Education and some co-developments (software/hardware) performed in collaboration with private partners in WP3, to support exploitation. The reasoning for these will be duly justified, as per the "as open as possible, as closed as necessary" principle.

3.2 Making Data Accessible

Edu4ClimAte will generate and have access to a wide breadth of data, including observational and simulation data. As a general principle data will be made openly available, and Edu4Climate will ensure that they are as open and accessible as possible, given that data produced through consortium activities is not bound by contractual obligations or other form of confidentiality restrictions.

Repository: Data will be deposited in trusted repositories, which are widely used by the project's thematic community. Consortium Partners are encouraged to used Zenodo and EBAS for Edu4Climate purposes, as it is well-established in the atmosphere and climate sciences domain, though to avoid any complications with Beneficiaries' institutional or other guidelines, any trusted repository (i.e. a certified or trusted repository which meets the necessary criteria under Horizon Europe, including ensuring that data is assigned a unique identifier), may be used. For the purposes of also maximizing visibility and accessibility to publications and other research outputs, Beneficiary Open Institutional repositories will also be used (e.g. the Coordinator's open access repository (https://repository.cyi.ac.cy/).

Data: In line with Open Science principles, Edu4Climate will adhere by as open as possible as closed as necessary. Whilst by default project data will be made openly available, specific cases where exceptions need to apply have been identified, whereas certain datasets cannot be shared or need to be shared under restricted access conditions. The identified exceptions relate to cases where data:

- In case embargo is applied to allow time to publish (e.g. in scientific journals) or seek protection of the intellectual property (e.g. in the case of registered IP protection applications such as patents).
- In case opening the data goes against the legitimate interests of specific beneficiaries (e.g. in case of interest for commercial exploitation) or other constraints as per the Grant Agreement (e.g. relating to sensitive data or other type of proprietary information).
- For Raw data (e.g. atmospheric model simulation data) will only be made openly available after it has been post-processed, has gone through appropriate quality control and has been published in an official capacity (e.g. in a scientific journal), to mitigate any risks or conflicts of interest arising

[1] EBAS (database hosting observation data of atmospheric chemical composition and physical properties), http://ebas.nilu.no/

from the premature sharing of ongoing work. The Consortium will leverage deposition repositories on the servers of the Coordinator, and provide instructions on how to be accessed (including a validation process to ascertain the identity of the person requesting to access the data).

More specifically, to support the implementation of the above, Edu4Climate scientific data and other relevant research outputs are classified into three categories in regards to accessibility: **i)** openly/publicly accessible (O), **ii)** private (P), and **iii)** sensitive/confidential/restricted (R) data. In further detail:

i) Openly/publicly accessible data - O

These are data that are accessible to all, free of charge, and can be accessed online in the public domain. Once these data have been made openly and publicly accessible, they might be modified or disclosed unauthorized, but any such misuse of data placed in the open data category should pose no or only minimal risk impact to the activities of Edu4Climate. Though, data curators will retain a level of monitoring, to the extend feasible, as a precautionary measure. Examples of (O) data may be:

- Research publications and technical reports published in Open Access journals and journals that promote the publication of research data in open access repositories (e.g., Zenodo/OpenAIRE).
- Non-processed measurements of atmospheric variables at ground level to be made available via the Boost Project 3 and citizen science (Task 5.3) activities (public signpost and free mobile app).
- Post processed measurements of atmospheric variables to be available via EBAS a freely accessible environmental repository of high standards. As soon as atmospheric variables related to other networks, such as ICOS, will be produced they will be included into their respective databases.

ii) Private data – P

These data need to be accessible by specific stakeholder groups, such as Project Beneficiaries or other partners beyond the Consortium in scientific collaborative projects, stemming from Edu4Climate dissemination and exploitation and/or complimentary to its activities, but are not meant to be publicly accessible. Any dataset not categorized either as public or restricted is automatically classified as private. More generally, data/information will be classified as private (P) when their unauthorized disclosure and/or modification may result in a medium risk impact on the ongoing research activities of CARE-C and its advanced partners, and for this reason, a higher level of security regarding their access/dissemination will be enforced. However, these data may become publicly available later, following an embargo period. The embargo lifting may be decided in regards to the timeframe and with the terms that the data curator(s) is required to follow or decides best to do so. The aim is always to provide publicly available (O), high-quality/verified research data relevant to Edu4Climate activities, in line with Open Science. For instance, data of this type (P) may include:

- Scientific code created to perform specialized data analysis, before its official publication (following a successful peer-review evaluation).
- Source code under development that needs to be tested and verified.
- Atmospheric measurements that are collected as part of a collaborative project either among Consortium partners or with institutions outside the consortium. Sharing data under this category may jeopardize the collaboration.

iii) Sensitive, confidential or restricted data – R

These data will only be accessible to members of the Edu4Climate Consortium (via private email or the Edu4Climate private project management system) and made available to the European Commission (via the EC Portal) and any other required third parties only when strictly necessary to adhere to regulatory demands or for business continuity purposes. In general, data will be identified as restricted whenever any unauthorized disclosure and/or modification could jeopardize consortium beneficiaries' interests. Therefore, the highest level measures of access control will be imposed on sensitive information and restricted data. As is the case with private data, restricted (R) data may be placed under an embargo for a certain amount of time until the person(s) responsible for the data determines that the restriction is no longer pertinent and/or when the (R) data have gone through the appropriate scientific evaluation and verification. Data under the (R) category include:

- Those related to ongoing research projects that have not gone through a peer-review evaluation, or that may inflict conflict of interest for exploitation, or that involve IP restrictions.
- Data bound by contractual obligations or confidentiality restrictions, or which contain sensitive information.

In all cases, the reason for why data cannot be made openly available immediately will be justified, and how long this will apply will be determined, to ensure that research data are made available as soon as possible. As Edu4Climate results emerge, the Consortium will further define and implement the process(s) through which data will be accessed, signpost these on the Project website, and report on them accordingly.

The data will be accessible through a free and standardized access protocol. The identity of the person accessing the data will be ascertained through registration.

Metadata: Metadata will be made openly available and licensed under a public domain dedication CC0, as per the Edu4Climate Grant Agreement. Reference to the software through which the data can be accessed or read will be included, prioritizing open applications wherever possible. Metadata will contain information to enable the user to access the data. The data and metadata will be downloadable together.

Data Access Committee: To better implement the above and to set further underlying guidelines regarding data access and availability as needed, the Consortium is forming a Data Access Committee (DAC). Within its remit, the Committee will also handle special situations that involve "sensitive" data, for instance produced in collaboration with private partners (WP3). To ensure coordination and coverage of the needs and requirements across the Project Work Programme, the DAC will involve the participation from the leaders of all Edu4Climate Work Packages, which bear responsibility for communicating and ensuring satisfactory input and implementation of Data Management as per the DMP, in regards to their respective WP activities.

3.3 Making Data Interoperable

A large volume of Edu4ClimAte data, for example climate model data outputs or data collected in the context of Atmospheric Research Facilities which follow the EBAS protocol, is already interoperable (e.g. in NASA AMES or NetCDF) and can be accessed and processed by relevant open and/or commercial software applications. Overall, Edu4climate will strive to follow community-endorsed interoperability best practices, aligned with the data and metadata vocabularies, standards, formats and methodologies followed by the European Environmental Research Infrastructure (ENVRI) community, and the latest community good-practice guidelines developed through the H2020 ENVRI-FAIR project, which aimed 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.

In line with the Horizon Europe Model Grant Agreement, for restricted or otherwise confidentially designated data, data interoperability, exchange and reuse will be restricted to members of the Edu4ClimAte Consortium, European Commission services and government agencies of the Widening Countries (e.g. DLI) where relevant. In such cases, for purposes of internal interoperability and reusability and for quality assurance, all consortium partners are required to use English language in all communications, and to use widely available documentation and management tools and software (e.g. MS Office).

3.4 Increase data re-use

To permit the widest re-use possible, publicized data that have gone through the appropriate processing and quality control, will be made available in the public domain as soon as possible, for use by third parties in non-commercial research, educational and other wider knowledge exchange purposes. The intention is for this data to be made and remain available for indefinite use, or for as long as they remain relevant, even after the end of the project. As per the obligations set out in the Edu4Climate Grant Agreement, the use of DOIs, metadata and CC-BY License will be used to support this, and utilizing trusted repositories such as Zenodo and the European Open Science Cloud. As part of this process, the provenance of the data will be thoroughly documented using the appropriate standards. In accordance with the Edu4ClimAte project's IPR strategy, data and data outputs

produced will be classified using levels and criteria defined according to final use, and regulated to manage ownership and access to key knowledge. The intent is to follow the EBAS established format, whereas documentation needed to validate data analysis and facilitate data re-use is included in the metadata. For additional clarification the user maybe directed to the related publication(s).

4. 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 Edu4Climate, either digital (e.g. a mobile app or other web application to be developed as part of WP5, as well as any other software, workflows, protocols, etc.) 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 Edu4Climate periodic reviews.

5. 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 project (as per the project budget breakdown and in line with the Edu4Climate Grant Agreement conditions), as and where eligible (i.e. for fully-open access publications) and by 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 through the use of the established databases and trusted repositories mentioned above (e.g. via upload to EBAS).

6. Data Security

CyI will assume responsibility to ensure that project data are stored securely, and provisions are made to ensure that data security extends beyond the lifespan of the project. Data will be safely stored in trusted repositories (such as Zenodo) for long term preservation and curation. Whilst more specifically in regards to data recovery and secure storage/archiving, CyI will assume responsibility to ensure that project data are stored securely, and provisions are made to ensure that data security extends beyond the lifespan of the project. Specifically, the measures taken to ensure optimal data security include:

- limited writing access
- account control system
- working copies outside the primary data storage (for instance using cloud services)
- data transmission through a dedicated VPN infrastructure and stored in high availability databases (hosted within CyI premises)
- multiple compute and storage nodes

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

7. Ethics

Edu4ClimAte 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 Data Protection Officer (DPO) as well as involve its externally assigned collaborator for Data Protection matters. These measures will allow full compliance with data privacy rules, ensuring appropriate stewardship and curation of all data and research materials, while appropriately acquiring informed consent.

The Consortium will also comply 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 Edu4limate 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 of the above will be coordinated as horizontal for the Project via the dedicated Task 6.3 Ethics (Lead: CyI, Partners: All) (M1-M48).