



Horizon Europe Twinning Call (HORIZON-WIDERA-2023-ACCESS-02)	
PROJECT TITLE	Atmospheric and Solar Research and Innovation in the Eastern Mediterranean
PROJECT ACRONYM	ATARRI
Work Package	WP1 – Coordination and Management I
Deliverable	(D2) D1.2– Data Management Plan



Funded by the European Union



ACKNOWLEDGEMENT: This project has received funding from the European Union’s Horizon Europe Framework Programme (HORIZON-WIDERA-2023-ACCESS-02, Twinning Call) under the grant agreement No 101160258. DISCLAIMER: 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. Neither the European Union nor the granting authority

 Funded by the European Union	<p>This project is funded under the European Union’s Horizon Europe Twinning Call (HORIZON-WIDERA-2023-ACCESS-02) under grant agreement No 101160258</p>	
Call:	HORIZON-WIDERA-2023-ACCESS-02	
Project full title:	Atmospheric and Solar Research and Innovation in the Eastern Mediterranean	
Project acronym:	ATARRI	
Work Package (WP):	WP1 – Coordination and Management I	
Deliverable (D):	(D2) D1.2– Data Management Plan	
Due date of deliverable:	31 March 2025 (Month 6 of the project)	Version: 1.0
Author(s):	Rodanthi-Elisavet Mamouri, Argyro Nisantzi, Francesco Scarlatti (ECoE); Emanuele Emili, Tito Vintimilla Cruz, Francesco Benincasa (BSC-CNS); Anton Lopatin (GRASP SAS); Lionel Menard (ARMINES); Rodrigo Amaroe Silva (FC.ID); Kyriaki Papachristiopolou, Stelios Kazadzis (PMOD/WRC)	
Start date of project:	01/10/2024	Duration: 36 months

can be held responsible for them. © Copyright by the ATARRI consortium, 2023-2027. More info regarding the project you can find here:

<https://atarri.eu>

Dissemination Level		
PU	Public	✓
SEN	Sensitive, limited under the conditions of the Grant Agreement	✓

Document Sign-off		
Status of the document	Organization	Date
Drafted	ECoE	January 2025
Reviewed	ALL	5 March 2025
Approved	ALL	24 Month 2025
Submitted	ECoE	31 Month 2025

List of Tables

Table 1: Types of data and origin..... 7
Table 2: Preliminary list of types of data and origin for the implementation of the Exploratory Research Projects of WP8
..... 10

List of Figures

No table of figures entries found.

List of Acronyms and Abbreviations

ARMINES	ASSOCIATION POUR LA RECHERCHE ET LE DEVELOPPEMENT DES METHODES ET PROCESSUS INDUSTRIELS
BSC CNS	BARCELONA SUPERCOMPUTING CENTER CENTRO NACIONAL DE SUPERCOMPUTACION
CIENCIAS UL	FACULDADE DE CIENCIAS DA UNIVERSIDADE DE LISBOA
EC	European Commission
ECoE	ERATOSTHENES Centre of Excellence
EU	European Union
FC.ID	FCIENCIAS.ID-ASSOCIACAO PARA A INVESTIGACAO E DESE
GRASP SAS	GENERALIZED RETRIEVAL OF ATMOSPHERE AND SURFACE PROPERTIES EN ABREGE GRASP
PMOD WRC	SCHWEIZERISCHES FORSCHUNGSINSTITUT FUER HOCHGEBIRGSKLIMA UND MEDIZIN IN DAVOS
WP	Work Package

Summary

This document is prepared for the purposes of “D1.2– Data Management Plan” of the project ATARRI. The Project is funded by Horizon Europe Framework Programme (HORIZON) - Coordination and Support Actions under grant agreement no. 101160258. The ATARRI project aims to enhance Earth Observation R&I and modelling capacities of the Eratosthenes Centre of Excellence (ECoE) with the goal to create a future center for frontier climate and weather research and services in the Eastern Mediterranean.

The main aim of this document is to provide insights about the plan of data management of the ATARRI project, namely, how the data and metadata managed inside the project have been or are intended to be align with the FAIR principles. Along with this useful information the data themselves used or intended to be used in the project are reported in tabular form with their origin, the related working package where they are involved and the partner of the ECoE in the project that will help ECoE to gain new insights about its scientific objective in ATARRI.

Table of Contents

LIST OF TABLES	5
LIST OF FIGURES	5
LIST OF ACRONYMS AND ABBREVIATIONS.....	5
SUMMARY	6
1. INTRODUCTION.....	7
2. DATA SUMMARY	7
3. FAIR DATA.....	12
3.1. <i>Making data findable, including provisions for metadata</i>	<i>12</i>
3.2. <i>Making data accessible.....</i>	<i>12</i>
3.3. <i>Making data interoperable</i>	<i>13</i>
3.4. <i>Increase data re-use.....</i>	<i>13</i>
4. OTHER RESEARCH OUTPUTS	14
5. ALLOCATION OF RESOURCES.....	15
6. DATA SECURITY.....	15
7. ETHICS.....	15
8. OTHER ISSUES	15

1. Introduction

The Data Management Plan (DMP) of ATARRI project provides guidelines for the management of the research data generated within the project, including the collection, handling, quality control procedures, archiving and accessibility to the data, during and after the end of the project. The plan is based on the **Horizon 2021 DMP template**, it will follow the **FAIR principles for data** (Findable, Accessible, Interoperable, Reusable), and it will be regularly updated during the project.

The DMP is expected to be used as a baseline for a more detailed and extended plan for managing the observations acquired at CARO for Earth observation-based studies and higher-level products retrieved by ATARRI modeling analysis in the future.

2. Data summary

This section includes the types of data to be collected, generated and re-used within the ATARRI project. Since these activities have just started and are still ongoing, the summary provided here will be only indicative of the data that is expected to be used and generated within the project. An updated list will be provided in the updated versions of the DMP.

These data will include:

- **Ground-based remote sensing datasets** used in WP3 and WP4, for the purposes of training activities foreseen in the project and respective publications. This dataset will be mainly acquired from the CARO observations in Limassol.
- **Synergistic Datasets** used in WP8, for the research demonstration study. This data will be generated by ECoE in collaboration with the partners.
- **Ground-based and satellite remote sensing data** used in WP3, WP4, WP8, for the training activities and the exploratory research project . These data will be acquired from CARO and from Space Agencies.
- **Other research outputs** generated within the project.

The types and origin of the data are provided in the following Tables 1 and 2.

Table 1: Types of data and origin

Type of data and origin	List of data
ERATOSTHENES CoE Ground based datasets used for the implementation of the ATARRI project	
Aerosol and Cloud Ground-based remote sensing data Origin: ECoE/CARO CARO remote sensing aerosol and cloud platforms. The data will be available under ACTRIS and CLOUDNET repositories.	<ul style="list-style-type: none"> ● PollyXT lidar data: Aerosol Backscatter Coefficient ($m^{-1} sr^{-1}$) at 1064nm, 532nm and 355nm, Particle Linear Depolarization Ratio (PLDR) at 532nm and 355nm, Aerosol Extinction Coefficient (m^{-1}) at 1064nm, 532nm and 355nm, Lidar Ratio at 1064nm, 532nm and 355nm, Volume Linear Depolarization Ratio (VLDR) at 532nm and 355nm.

	<ul style="list-style-type: none"> • CIMEL sun-photometer products, provided by AERONET (list of products and data access at: https://aeronet.gsfc.nasa.gov/cgi-bin/data_display_aod_v3?site=Mindelo_OSCM&nachal=2&level=3&place_code=10). • MIRA35 Doppler Cloud Radar (METEK) data: Radar Reflectivity (Z), Doppler Velocity (m/s), Spectral Width (m/s), Linear Depolarization Ratio. • CHM15K Ceilometer (LUFFT) data: Backscatter Coefficient ($\text{m}^{-1} \text{sr}^{-1}$) at 1064nm, Aerosol layer in PBL (m), Cloud Base Height (m), Cloud Depth (m). • HATPRO-G5 Microwave Radiometer (Radiometer Physics) data: Brightness Temperature (K), Integrated Water Vapor (kg/m^2), Liquid Water Path (g/m^2), Vertical Temperature Profile (K), Vertical Absolute Humidity Profile (g/m^3), Meteorological Parameters Pressure, Temperature, Relative Humidity at 3m. • StreamLine XR Doppler Lidar (Halo Photonics) data: Horizontal Wind Speed (m/s), Horizontal Wind Direction ($^{\circ}$), Vertical Wind Velocity (m/s). • Parsivel2 Disdrometer (OTT Hydromet) data: Raindrop Size Distribution (mm), Precipitation Rate (mm/h), Drop Velocity (m/s), Kinetic Energy of Falling Particles (kJ), Number of Falling Particles, Equivalent Radar Reflectivity (dBZ).
<p>Solar radiation observatoions Origin: ECoE/Solar radiation stations The data will be available on Spatial Data Infrastructure and under a doi in the zenodo repository</p>	<ul style="list-style-type: none"> • 2 Pyranometers (EKO): Global Horizontal Irradiance (W/m^2, temporal resolution 1 min) and Diffuse Horizontal Irradiance (W/m^2, temporal resolution 1 min). • SunTracker (EKO). • Pyrhemometer (EKO): Direct Normal Irradiance (W/m^2, temporal resolution 1 min). • Erythemal UV actinometer (Kipp & Zonen) (W/m^2, temporal resolution 1 min). • Spectrophotometer (DMc150 double monochromator, Bentham) (Spectral Solar irradiance (spectral range from ~280 to 600nm).
Datasets beyond ECoE/CARO used for the training in WP3 and WP4	
BSC-CNS trainings	
<p>MPLNET Lidar data containing aerosol properties from a selected period of observation. Origin: NASA-MPLNET</p> <p>Sun photometer data Origin: NASA-AERONET</p>	<p>MPLNET LiDAR data:</p> <ul style="list-style-type: none"> • Aerosol optical depth • Aerosol backscatter coefficient • Depolarisation ratio • Extinction coefficient <p>Aerosol observations (O’Neill and Direct Sun algorithms at Level 1.5):</p>

	<ul style="list-style-type: none"> • Aerosol optical depth • Angstrom exponent
Data containing satellite observations for aerosol properties Origin: VIIRS-NOAA	Satellite observations at Level 2 v2: <ul style="list-style-type: none"> • Aerosol optical thickness (Land-Ocean best estimate) Aerosol type (Land ocean)
Dust forecast model outputs Origin: BSC CNS WMO-SDS WAS	MONARCH dust forecast operational data: <ul style="list-style-type: none"> • Dust aerosol optical depth (AOD) • Dust concentration • Dust load
GRASP SAS trainings	
Raw lidar and photometers data Origin: GRASP/ECOE	<ul style="list-style-type: none"> • Lidar signal profiles as provided from the lidar simulator for ground-based systems
Inversion model outputs Origin: GRASP/ECOE Data containing aerosol properties obtained from inversion of raw sunsphotometr and lidar data for selected period of observation	Aerosol columnar properties (for one or several modes): <ul style="list-style-type: none"> • Size distribution. • Chemical composition and/or complex refractive index. • Fraction of dust particles. • Spectral AOD. • Spectral SSA. • Spectral Lidar ratios. • Spectral Phase functons. Aerosol vertical properties (for one or several modes): <ul style="list-style-type: none"> • Aerosol vertical concentration profile. • Profile of spectral extinction. • Profile of spectral backscatter. • Profile of spectral lidar ratio. • Profile of spectral SSA.
ARMINES/ENSMP Trainings	
Urban descriptive data Origin: Department of Land and Survey, Republic of Cyprus	<ul style="list-style-type: none"> • Digital surface model. • Partial Limassol data available for 2019
Urban shadowing model outputs Origin: SAFaas software	<ul style="list-style-type: none"> • Urban shading estimates.
PMOD/WRC Trainings	
Radiative Transfer model outputs Origin: PMOD/WRC	<ul style="list-style-type: none"> • Vertically resolved radiometric quantities and heating rates outputs as provided from radiative transfer model (libRadtran package) using ground- and satellite- based data input.

FC.ID/CIENCIAS UL Trainings	
Ground based aerosol, cloud, radiation data Origin: CARO ECoE	<ul style="list-style-type: none"> CARO ECoE aerosol observations. In-situ radiation data.
Urban descriptive data Origin: Google Earth Engine; Department of Land and Survey, Republic of Cyprus	<ul style="list-style-type: none"> Digital surface model. Cyprus Surface model Partial for Limassol data available for 2019
Regional Radiation and aerosol data Origin: CAMS	<ul style="list-style-type: none"> Radiation data. Air quality data
Weather input data Origin: ERA5	<ul style="list-style-type: none"> Reanalysis-derived air temperature data.

Table 2: Preliminary list of data and origin for the implementation of the Exploratory Research Project of WP8

Remote sensing data from CARO and satellites for the exploratory research project in WP8	
Type of data and origin	List of data
BSC CNS	
Dust forecast model outputs Origin: BSC - MONARCH	<ul style="list-style-type: none"> Tailored dust forecast products.
GRASP SAS	
Input and output data of GRASP model Origin: GRASP	<p>Aerosol columnar properties (for one or several modes):</p> <ul style="list-style-type: none"> Size distribution. Chemical composition and/or complex refractive index. Fraction of dust particles. Spectral AOD. Spectral SSA. Spectral Lidar ratios. Spectral Phase functions. <p>Aerosol vertical properties (for one or several modes):</p> <ul style="list-style-type: none"> Aerosol vertical concentration profile. Profile of spectral extinction. Profile of spectral backscatter. Profile of spectral SSA. Profile of spectral lidar ratio. <p>Satellites data:</p>

	<ul style="list-style-type: none"> • Spectral AOD. • surface spectral BRDF parameters.
ARMINES/ENSMP	
Modelled data Origin: Google Earth Engine Origin: SAFaas software	<ul style="list-style-type: none"> • Digital surface model. • Urban shading estimates.
PMOD/WRC	
Radiative Transfer model outputs Origin: libRadtran/CARO	Radiative transfer model (libRadtran package) outputs using ground- and satellite- based data input obtained by CARO under strong dust intrusions to investigate the impact of strong dust events on spaceborne observations.
FC.ID/CIENCIAS UL	
Cyprus data and Origin: Department of Land and Survey	<ul style="list-style-type: none"> • Cyprus Digital surface model. • Photovoltaic timeseries • PVLIB using Python
Other research data generated/used in the project	
<i>To be provided in Updated version of the Data Management Plan and Final version of the Data Management Plan</i>	

Purpose of data in relation to the objectives of the project

All data will address the objectives of the project regarding strengthening the expertise and improving the research profile of ECoE/CARO, along with spreading excellence at the national R&I ecosystem, through strategic collaborations with experienced partners in the area of remote sensing and modeling applications.

Types and formats of data, along with their expected size

In order to ensure their interoperability, the data will be provided in commonly used formats (e.g., NetCDF, ASCII), which offer self-describing capabilities and platform independence. The expected size of the data per file will range from a few KB (e.g. AERONET products provided in ASCII files) to a few hundred MB (e.g. satellite data and lidar/radar data).

Data utility outside the ATARRI project

ATARRI will provide open-access to quality-assured datasets for weather and climate research, and model assimilation services. The data are expected to be exploited by the following users:

- Space Agencies (e.g., ESA, NASA)
- Academia (universities and research institutes)
- ESFRI projects (e.g ACTRIS, ICOS)
- Copernicus
- Regional stakeholders (e.g. Republic of Cyprus, Government, Regions)
- The industry (local SMEs and Space Sector)
- Society as a whole

3. FAIR data

The DMP of ATARRI will follow the **FAIR principles** for data, for making the data **Findable, Accessible, Interoperable**, and Reusable, and enhance their dissemination and exploitation during and after the end of the project.

3.1. Making data findable, including provisions for metadata

All data will be deposited in the trusted open access repositories of the ACTRIS – CLOUDNET Datacenters (<https://cloudnet.fmi.fi/search/visualizations?site=limassol>) for the aerosol and cloud CARO observations and Spatial Data Infrastructure (<https://viewer.webservice-energy.org/in-situ/#>) for Solar Radiation Station data and metadata, in addition the and Zenodo repository will be used to assigned a Digital Object Identifier (DOI) in order to make content easily findable and uniquely citable. Additionally, in the ATARRI website (<https://atarri.eu/>), a dedicated tab (under Material-> Datasets) will provide a summary of the datasets used, with short description and links to the aforementioned dataset repositories.

The data will be provided with respective **metadata**, including information at least about the following:

- Datasets: description, date of deposit, author(s), venue
- Licensing terms
- Persistent identifiers for the dataset (DOI)
- Authors involved in the action and their organizations
- Data keywords (for example: *Remote sensing, Ground-based, Satellite, Simulation, Training, Lidar, Radar, sun-photometer, Radiation*)

3.2. Making data accessible

Following the HORIZON open research data principles, the data and respective metadata of the ATARRI project will be accessible after their production, with exemptions following the principle “as open as possible, as closed as necessary”, as described in Sect. 3.2.1 below.

For scientific results to be published, Immediate Open Access will be chosen, and they will be deposited in the trusted open access repository of Zenodo and ACTRIS, Zenodo for Solar Radiation Station data and metadata. Upon depositing, a DOI will be automatically assigned. More specifically, a link to the ACTRIS will be established, in order to connect CARO to this long-term repository in Europe, for archiving and exchanging correlative data for validation of atmospheric composition products from satellite platforms. This link will be facilitated through our collaboration with ACTRIS central data facilities (for aerosol and cloud remote sensing data) and will make the CARO dataset accessible to the broader community.

The datasets that will be used in the framework of the ATARRI training and that will be produced by ECoE for the needs of the project will be deposited in the Zenodo repository (<https://zenodo.org/>) which is a general-purpose trusted open access repository with DOI, as this is defined in the ANNEX 5 of the HORIZON Annotated Model Grant Agreement ([aga_en.pdf \(europa.eu\)](#)).

Data and metadata license

The data and respective metadata will be published under the latest version of a Creative Commons Attribution International Public Licence (i.e., CC BY 4.024). This license follows the HORIZON open access to research data principle ‘as open as possible, as closed as necessary’, permitting the data to be used for any kind of activity, as long as the source is attributed, and the results are also published under the same license.

Although currently no exceptions are foreseen and all data used or generated throughout the project is planned to be publicly available, in case there are certain data that need to be kept confidential, justified reasons will be provided in the updated versions of the DMP.

3.3. Making data interoperable

Since the data production and collection is still in progress, their interoperability is subject to change. For the data in NetCDF (Network Common Data Form), the self-describing, machine-independent format will facilitate the exchange and reuse. Regarding the metadata characteristics that will ensure the interoperability of the data, these are provided in Sect. 3.1. For the naming of the parameters (i.e., nomenclature of the atmospheric variables), the ACTRIS RI naming conventions will be adapted, which are compliant with the Climate and Forecast conventions (i.e., [ACTRIS \(nilu.no\)](http://actris.nilu.no)).

For the Solar Radiation Station of Limassol the naming convention is obtained following the BSRN recommendations ([BSRN recommended QC tests](http://www.bsrn.org/)) and again the Climate and Forecast Conventions. The metadata and NetCDF naming convention used can be found on [Attribute Convention for Data Discovery 1-3 - Earth Science Information Partners \(ESIP\)](http://www.esipfed.org/docs/Attribute_Convention_for_Data_Discovery_1-3_-_Earth_Science_Information_Partners_(ESIP).pdf). The workflow used for the interoperability in the Solar Radiation Station followed the workflow: from not interoperable csv data as collected by the instrument computer Libinsitu tool (<https://libinsitu.org/>) has been used to convert the data into NetCDF format following the standard maintained by the Open Geospatial Consortium and following, as said, the Climate Forecast Convention for the naming vocabulary.

3.4. Increase data re-use

In order to increase data re-use, the data and their respective metadata will be made freely available immediately after their publication and will be disseminated through the dedicated page in the ATARRI website. The data and metadata will be licensed under the latest version of a Creative Commons Attribution International Public Licence (i.e., CC BY 4.024) in line with the obligations set out in the Grant Agreement. This way, the data produced in the project will be usable by third parties, even after the end of the project, available for at least the lifetime of the Zenodo repository. Indicatively, according to Zenodo's general policies (<http://about.zenodo.org/policies/>) "items will be retained for the lifetime of the repository ... which is defined for the next 20 years at least". The same Zenodo will be used to provide a repository of data and metadata from the Solar Radiation Station in Limassol.

Data documentation

The data will be accompanied by sufficient documentation of the measurements to have in-depth information on the quality of the data and will be properly labeled. Since the data production/collection is still in progress, the details included in the documentation will be provided in the updated versions of the DMP.

Standards and quality information

The data used and produced throughout the project will follow the ACTRIS and the Quality Assurance Framework for Earth Observation (QA4EO) recommendations on data standards and quality information. The **data standards** refer to how the data is collected, analyzed and distributed, including measurement protocols,

quality assurance procedures, quality control, quality assessment metadata standards. The **data quality information** refers to the reliability of the data and is a factor for defining their suitability for their exploitation, facilitating their interoperability and reuse, based on the FAIR data principles.

ACTRIS recommendations

A number of quality standards and quality-assurance (QA) procedures are established within ACTRIS, in order to assure the high quality aerosol and cloud remote sensing datasets, (e.g., for aerosols: <http://actris.nilu.no/Content/?pageid=13d5615569b04814a6483f13bea96986>). Standard handling of the data:

1. The data are processed by common standardized automatic analysis software developed within ACTRIS.
2. The produced products follow the recommendations and quality guidelines defined by ACTRIS, and are accompanied by ACTRIS-defined standards and metadata. Moreover, information required to precisely describe the measurements, as variable standardization and characterization of data uncertainty, are also considered.

Within the ATARRI project, additional internal quality screening procedures will be performed, for each dataset used.

QA4EO recommendations

The QA4EO guidelines aim at facilitating interoperability and harmonization of data and derived products, by e.g. using fully traceable Quality Indicators, enabling users to readily assess the fitness for the purpose of the data. Moreover, the QA4EO framework includes a set of general guides to establish reference standards, perform comparisons (of the instruments and methods), develop validated models and algorithms, provide the uncertainty of the measurements, and establish quantitative evidence of traceability.

The data used in the ATARRI project will (in their majority) include information on the calibration of the instruments to appropriate references, the description of any auxiliary information brought into the processing of level 0 or 1 data products, as well as the characterization of the sensitivities due to the environmental conditions. Additionally, uncertainty analysis of all sources of uncertainty and the propagation of that uncertainty through the traceability chain is foreseen for most of the ATARRI datasets.

BSRN recommendations

For the QA and Quality control (QC) tests of the data provided by the Solar Radiation Station of Limassol the BSRN guidelines have been used ([BSRN recommended QC tests](#)) to ensure high quality of the solar remote sensing that will be stored continuously in a Zenodo repository.

4. Other research outputs

Throughout the ATARRI project, we may need to re-use complementary datasets, models and software, provided to the consortium from third parties. In this case, we will examine the possibility of providing them for re-use, in line with the FAIR principles (e.g. in the online repositories), according to their respective licenses. Relevant information will be provided in the updated DMP.

5. Allocation of resources

The deposit of the data and metadata in the trusted open access repositories such of Zenodo is free of charge. The persons responsible for the data management, will be the Project Coordination Team (i.e., R odanthi-Elisavet Mamouri PI and Argyro Nisantzi PM of ATARRI), and the WP leaders. These people will coordinate with the rest of the researchers of the project, for the seamless management of the data throughout the project, including the collection, handling, quality control, archiving and providing access to the data.

6. Data security

The long-term preservation of the data and metadata is ensured by their deposit in the trusted open access repositories as Zenodo (as discussed in Sect. 3.4). The datasets will be also stored in ECoE servers, to ensure their preservation.

7. Ethics

All ethical parameters related to the data used and generated within the project will be provided in the Ethics page in SIGMA.

8. Other issues

No other national/funder/sectorial/departmental procedures for data management are identified. Updates will be provided in the updated versions of the DMP.