



**Co-funded by  
the European Union**



**Horizon Europe  
(HORIZON-CL5-2021-D1-01)**

# **Non-CO2 Forcers and their Climate, Weather, Air Quality and Health Impacts**



**Deliverable 8.7**

**Data Management Plan**

Grant Agreement No.	101056783
Project acronym	FOCI
Project full title	Non-CO2 Forcers and their Climate, Weather, Air Quality and Health Impacts
Call	HORIZON-CL5-2021-D1-01
Deliverable name	D8.7 Data Management Plan
WP contributing to the deliverable	WP8
Task producing the deliverable	Task 8.3
Type	<input type="checkbox"/> Report
	<input type="checkbox"/> Prototype
	<input type="checkbox"/> Demonstrator
	<input checked="" type="checkbox"/> Other
Dissemination level	<input checked="" type="checkbox"/> Public
	<input type="checkbox"/> Sensitive
	<input type="checkbox"/> UE/EU-Restricted
Due date of deliverable	Month 6
Actual submission date	Month 6
Lead beneficiary	ECMWF
Author(s)	Richard Engelen (ECMWF), Ranjeet S Sokhi (UH), Tomáš Halenka (CU), Helena Nyklová (CU), Alexander Baklanov (WMO), Leilani Dulguerov (WMO), Rosa von Borries (WMO), Christian Schweizer (WHO), David Grawe (UHam), Marco Pandolfi (CSIC), Johan Kuylenstierna (SEI), Colin Price (TAU), Sandro Finardi (ARIANET)
Other Contributor(s)	
Reviewer(s)	Alberto Troccoli (WEMC)
Keywords	FOCI, deliverables, data management, open data

### ACKNOWLEDGEMENTS

This project has been co-funded by the European Union with funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101056783 and from UKRI under the UK Government's Horizon Europe Guarantee (UKRI Reference Numbers: 10040465, 10053814 and 10050799).

Version	Date	Modified by	Comments
1.0	27 Feb 2023	Richard Engelen for comments by WP8 team and internal reviewer	First order draft written
2.0	6 Mar 2023	Richard Engelen	Comments from internal review taken into account
3.0	8 Mar 2023	Richard Engelen	Comments from WP8 taken into account
4.0	14 Mar 2023	Richard Engelen & Ranjeet Sokhi	Further additions to document
5.0	22 Mar 2023	Richard Engelen	Comments from second internal review taken into account

	Name	Date
Verification Final Draft by WP leaders	Alexander Baklanov (WMO)	
Check before upload by project Coordinator	Tomas Halenka (CU)	

## TABLE OF CONTENTS

<b>TABLE OF CONTENTS</b> .....	<b>3</b>
<b>EXECUTIVE SUMMARY</b> .....	<b>4</b>
<b>1. INTRODUCTION</b> .....	<b>5</b>
<b>2. SCOPE AND OBJECTIVES OF THE DATA MANAGEMENT PLAN</b> .....	<b>6</b>
2.1 Scope of the Data Management Plan.....	6
2.2 Objectives of the Data Management Plan.....	6
<b>3. DATA SUMMARY</b> .....	<b>7</b>
3.1 Approach.....	7
3.2 Introduction to the data sets .....	7
3.3 Re-used datasets .....	10
3.4 Project generated datasets .....	12
<b>4. DATASETS TO SUPPORT STAKEHOLDERS</b> .....	<b>13</b>
4.1 Datasets to support policy and decision makers.....	13
4.2 Datasets to support of operational services .....	13
4.3 Intellectual Property Rights and data protection .....	14
4.4 Quality assurance and control management of datasets .....	14
<b>5. FAIR DATA</b> .....	<b>15</b>
5.1 Making data findable .....	15
5.2 Making data accessible .....	15
5.3 Making data operable.....	16
5.4 Increase data re-use.....	16
<b>6. ALLOCATION OF RESOURCES</b> .....	<b>16</b>
<b>7. CONCLUSIONS</b> .....	<b>16</b>
<b>8. REFERENCES</b> .....	<b>17</b>
<b>APPENDICES</b> .....	<b>17</b>
Appendix 1. List of Partners .....	17
Appendix 2. List of Abbreviations .....	18

## EXECUTIVE SUMMARY

This document is the deliverable “D8.7: Development of the Data Management Plan” for the European Union project “FOCI: Non-CO2 Forcers and their Climate, Weather, Air Quality and Health Impacts” (hereinafter also referred to as FOCI, project reference: 101056783).

This deliverable is written against the backdrop of the overall project objectives (i.e., improvement of the knowledge to address critical gaps in the understanding of the climate forcing of non-CO2 forcers) and outlines the data management plan, guiding the FOCI consortium partners and, as such, responding to the requirements of the H2020 and Horizon Europe Open Research Data Pilot, to document which research data will be re-used or produced by the FOCI project, in which format, and how it will be made available.

The document outlines the objectives of the Data Management Plan and provides a preliminary list of data sets used by Work Packages 1 to 7. This list will be periodically updated (with a frequency of at least six monthly) with the necessary details as they emerge during the implementation of the project, in consultation with the partners producing data.

This deliverable is primarily targeted at the consortium partners and should serve as a reference for the management of data products in the relevant deliverables. It also serves to support the cross-cutting activity on data integration and data products, which will interact with all WPs throughout the duration of the project to maximize benefits of the observational and modelling data generated by FOCI. All FOCI team members should refer to this plan to ensure a consistent approach to the sharing and documentation of the various datasets (including metadata). Further, in the later stages of the project, the updates and final version shall address broader external community to support dissemination and exploitation tasks in connection to CDE Plan.

## 1. INTRODUCTION

The FOCI project is funded by the European Commission under Call HORIZON-CL5-2021- D1-01-01: Improved understanding of greenhouse gas fluxes and radiative forcers, including carbon dioxide removal technologies, under grant agreement No 101056783. While the causes and global processes connected to well mixed greenhouse gases (GHGs) and their impacts on global to continental scales are well understood with a high level of confidence, there are knowledge gaps concerning the impact of many non-CO<sub>2</sub> radiative forcers and conclusions made have low confidence. FOCI aims to improve the knowledge to address critical gaps in the understanding of the climate forcing of non-CO<sub>2</sub> forcers and it targets significant scientific gaps identified in the IPCC AR6 (2021) and will undertake comparisons with the CMIP6 ensemble of models to evaluate the results.

To guide the FOCI partners with their management of used and produced datasets and in accordance with the Horizon Europe Model Grant Agreement, a data management plan ('DMP') is established and regularly updated. As per the Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020<sup>1</sup>, Research Data "Refers to information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form."

The Open Research Data Pilot "aims to improve and maximise access to and re-use of research data generated by Horizon 2020 projects" and applies to data sets that are "needed to validate the results presented in scientific publications". The Data Management Plan is expected to "specify what data will be open: detailing what data the project will generate, whether and how "it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved"<sup>2</sup>.

Within Task 8.3 (Development of the Data Management Plan), the project will develop and maintain this Data Management Plan that outlines the data management principles of the FOCI Project according to the Open Research Data Pilot of H2020 and describes the data management life cycle for all datasets to be collected, processed, and generated in the project. It constitutes the first version of the DMP document of the FOCI project and provides the baseline for the data management related activities for the whole consortium. More specifically, the Data Management Plan will cover the following activities:

- Types of data that will be collected (re-used) and/or generated.
- Types and formats of data that will be used.

---

<sup>1</sup> [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)

<sup>2</sup> [https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination\\_en.htm](https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm)

- The purpose of the data generation or re-use and its relation to the objectives of the project.
- The expected size of the data that the project intends to generate or re-use.
- The origin/provenance of the data, either generated or re-used.
- Users of the data ('data utility'), outside the project.

## **2. SCOPE AND OBJECTIVES OF THE DATA MANAGEMENT PLAN**

### **2.1 Scope of the Data Management Plan**

A Data Management Plan (DMP) is developed to set out the specifications for data, quality control, metadata generation, data access, data stewardship and how data will be maintained and preserved. As with scientific peer-reviewed publications, datasets generated by the project will be deposited in repositories and made Open Access. Data will be made freely available for use where possible. To facilitate the creation and monitoring of the Data Management Plan a specific Task 8.3 (WP8) is dedicated to this activity. A separate cross-cutting activity on data integration and data products will interact with all WPs throughout the duration of the project to maximize benefits of the observational and modelling data generated by FOCI. We will make data available and accessible according to agreed protocols, as defined in this DMP, to the wider community.

It is expected that a huge volume of data (including observational packages, model runs and post-processed model outputs) will be created during the project. In particular, new selected data products generated by the project to support model verification or related to specific process studies will be published with DOI and shared in public data repositories. Modelling results will be shared by relevant project partners accompanied by the extended metadata and descriptions that would allow for their further utilization outside of the project. Specific attention will be paid to the data formats to ensure that data can be broadly and easily used and postprocessed by the interested communities. The delivery of the datasets along with scientific publications, will be targeted to meet the timeliness of the IPCC and other major assessments cut-off dates to contribute to their development and hence maximise the impact of the project.

### **2.2 Objectives of the Data Management Plan**

To achieve the overall aim of the project, the DMP will adhere to the following objectives:

- (i) To develop and maintain a DMP that follows the data management principles of the project as was defined in the Open Research Data Pilot of H2020 and in Horizon Europe Grant Agreement Model.
- (ii) To provide guidance and a framework to integrate observational and modelling datasets and data products for improving and evaluating multiscale climate and atmospheric composition models

- (cross-cutting activity).
- (iii) To cooperate with Work Package leaders to oversee the generation, application, dissemination, and exploitation of data products arising from the project and maximise their impact.
  - (iv) To liaise with Work Package leaders to engage with stakeholders and other relevant projects who have an interest in the application and exploitation of datasets and products generated FOCI.
  - (v) To oversee issues related to Intellectual Property Rights associated with datasets and data products arising from the project.
  - (vi) To liaise with the QA/QC Committee to support the quality control process for datasets and data products before their dissemination and exploitation.
  - (vii) To manage the use of personal data.

### **3. DATA SUMMARY**

#### **3.1 Approach**

The Data Summary presented in this section provides an overview of all the data sets used and produced during the project. As such, it will be updated when new information becomes available. For this first version, a dedicated web-based tool (<https://researchers.ds-wizard.org>) was used to gather information from the FOCI project partners. The results from the tool have led to the information presented below. Throughout the project we will further update the information by using both the tool above as well as more targeted questionnaires to be sent to all project partners. The questionnaire will especially support the gathering of information about newly produced data sets, their metadata, and means for sharing, both within the consortium and with the external science and user communities. The DMP will be updated accordingly approximately every 6 months.

#### **3.2 Introduction to the data sets**

Within WP1/WP2 (Integrating observations), the project will make full use of existing observational infrastructures operated by other past and current projects, national organisations and global coordination efforts such as WMO/GAW. A list of these is given in Table 1. For analysis of the processes connected to the non-CO2 effects, observational data from different sources and campaigns are necessary. These observations will be integrated to address the key science objectives of FOCI.

**Table 1. List of major ground-based and satellite observational platforms to be used within FOCI**

Network/Project acronym	Network/Project name	Aerosol particle properties relevant to radiative forcing on climate	Relevant papers
GAW/ACTRIS	Global Atmosphere Watch/ Aerosol, Clouds and Trace Gases Research Infrastructure	Aerosol scattering, absorption, extinction coefficients and efficiencies, single scattering albedo (SSA) and other climate-relevant optical properties	Laj et al., (2020)
		Cloud condensation nucleus concentration (CCN)	
		Particle number concentration (CN)	
		Particle number size distribution (PNSD)	
		Particle mass (PM <sub>x</sub> ) concentration	
	Elemental and organic carbon (EC and OC) and black carbon (BC) concentration		
EARLINET/ACTRIS	European Aerosol Lidar Research Network/ Aerosol, Clouds and Trace Gases Research Infrastructure	Vertical profile of the particle light extinction and backscatter coefficients	Pappalardo et al., 2014
AERONET	Aerosol Robotic Network	Aerosol (extinction and absorption) optical depth (AOD) and SSA Column integrated aerosol size distribution	Schafer et al., 2019
COLOSSAL	Chemical On-Line cOmpoSition and Source Apportionment of fine aerosOL	Chemical composition of sub-micron aerosol particles chemical composition (OA, sulphate, nitrate, ammonium, chloride); POA and SOA sources.	Bressi et al., 2021
GASSP	Global Aerosol Synthesis and Science Project	CN, CCN, PNSD, PM concentration, PM chemical composition, and BC mass concentration from multiple campaigns.	Reddington et al., (2017)
WDCGG (data center under GAW)	World Data Centre for Greenhouse Gases	Mixing ratios of greenhouse gases (CO <sub>2</sub> , CH <sub>4</sub> , CFCs, N <sub>2</sub> O, etc.) and CO in the atmosphere	Schultz et al., 2015
<b>Instrument</b>	<b>Satellite</b>	<b>Data provider</b>	<b>Measured variables</b>
SLSTR	Sentinel-3	ESA/EUMETSAT	AOD
TROPOMI	Sentinel-5p	ESA	SO <sub>2</sub> , CH <sub>4</sub>
MODIS	Aqua, Terra	NASA	AOD
VIIRS	S-NPP	NOAA	AOD
PM <sub>Ap</sub>	Metop-A, -B, -C	EUMETSAT	AOD

Within WP3/WP4 (Multiscale modelling of the impact of non-CO<sub>2</sub> radiative forcers), FOCI will conduct analysis of the non-CO<sub>2</sub> radiative forcers and their impact with two state of the art Earth System Models (ESMs), five regional climate and atmospheric composition models and one high resolution microscale model



(see Table 2). All models will produce a significant amount of output data for analysis, exchange within the consortium, and in some cases for external use.

**Table 2. List of the main models that will be used within FOCI**

Model and partner	Atmosphere	Aerosols and atmospheric chemistry	Schemes
<b>ESM</b>			
EC-Earth3-AerChem Partners: BSC, KNMI	Modified version of the ECMWF Integrated Forecasting System (IFS) GCM, cycle 36r4; horizontal resolution TL255 (~80 km); 91 levels; model top at 0.01 hPa	TM5 for interactive tropospheric aerosols and atmospheric chemistry; stratospheric aerosols prescribed	Ocean: NEMO-LIM 3.6 ; ORCA1 (~ 1°); 75 levels Land: H-TESSSEL (IFS)
EC-Earth4 Partners: BSC, KNMI	Modified version of OpenIFS; current release based on IFS cycle 43r3; upgrade to 47r3 expected; horizontal resolution Tco95	Integrated in OpenIFS (see WP3)	Ocean: NEMO4-SI3; ORCA1 Land: H-TESSSEL
EMAC Partner: MPI-C	Spectral resolution of T63, i.e. 1.8x1.8 degree; 90 vertical levels; model top at 0.01 hPa	MECCA model for gas tracers (Sander et al. 2019), 7 log-normal modes for aerosols, GMXE submodel	Ocean: MPIOM Land: LPJ-GUESS
IFS Partner: ECMWF	Integrated Forecasting System (IFS)	CB05 for tropospheric chemistry and IFS-AER bulk scheme for aerosol.	<a href="https://confluence.ecmwf.int/display/CKB/CAMS%3A+Global+atmospheric+composition+forecast+data+documentation">https://confluence.ecmwf.int/display/CKB/CAMS%3A+Global+atmospheric+composition+forecast+data+documentation</a>
<b>Regional and urban scale models</b>			
RegCM Partner: CU, UH	Terrian -following sigma coordinate model with MM5 dynamical core	Simplified internal – on-line, RegCM-CAMx – off-line	Ocean: Climatological SST Land: CLM4.5
WRF-CMAQ Partner: UH, CU	Terrain-following non-hydrostatic- mesoscale model, resolution upto 1km	Bulk and sectional aerosol schemes (GOCART, SORGAM, MOSAIC)	Ocean: Climatological SST Land: NOAH LSP
Enviro- Partner: UHel	downscaling (regional-to-urban scales) 15-5-1+ km horiz. Resol., vertical levels – 40-60	aerosol microphysics M7, chemistry CBMZ	Ocean: SST, sea-ice data assimilation Land: GLCC ECOCLIMAP CORINE
FARM Partner: ARIANET	Terrain -following, offline coupled with meteorological driver (25-1 km)	SAPRC99 gas chemistry and AERO3 modal aerosol scheme	Land: CORINE or ESA CCI Land Cover
MITRAS Urban climate model Partner: UHam	Resolution of 1-15m	Non-reactive transport	Land: WUDAPT
<b>Analysis tools</b>			
LEAP Emission scenario analysis tool Partner: SEI			
Agricultural emissions and health assessment tool Partner: SEI			

In WP5/WP6 (Emissions and scenarios), FOCI will develop, for particular case study regions, countries and cities, more specific assessments of the implementation of non-CO<sub>2</sub> mitigation measures taking into account

national circumstances, policies, plans and development visions and isolating those from other mitigation actions. This development consists of adapting global emission scenarios, such as the Shared Socio-economic Pathways (SSPs), for particular national circumstances (political, social developmental) that impact the effectiveness of implementation of these mitigation measures. This development will result in specific emission data sets as well as a limited set of dedicated model simulations.

### 3.3 Re-used datasets

As part of this first version of the DMP, we have summarised the input datasets for which we gathered more detailed information than in the previous section. As the project progresses, more datasets will be added to the following descriptions.

- **ECMWF ERA5 reanalysis** (<https://cds.climate.copernicus.eu/>)

Owner of this dataset: ECMWF on behalf of European Commission, Hersbach et al (2020), doi: 10.1002/qj.3803.

The dataset can be used in the provided format without any conversion needed, although interpolation might be applied depending on the exact usage.

We will use version “ERA5” of this dataset. If a new version becomes available during the project, we will stay with this version. The provider keeps old versions around so the same reference data will be available to reproduce our results.

We will use the dataset as follows: In atmosphere-nudged Earth System Model simulations.

- **AerChemMIP CMIP6 model outputs** (<https://esgf-node.llnl.gov/search/cmip6/>)

Owner of this dataset: Publisher: Earth System Grid Federation.

The dataset can be used in the provided format without any conversion needed.

The provider keeps old versions around so the same reference data will be available to reproduce our results.

We will use the dataset as follows: Analysis of the influence of short-term climate forcings on climate.

- **CAMS European anthropogenic emissions – Air Pollutants** (<https://doi.org/10.24380/0vzb-a387>)

Owner of this dataset: ECMWF on behalf of European Commission; Contact details: TNO Department of Climate, Air and Sustainability – Utrecht – Netherlands – Hugo Denier Van der Gon ([hugo.deniervandergon@tno.nl](mailto:hugo.deniervandergon@tno.nl)) – Jeroen Kuenen ([10eroen.kuenen@tno.nl](mailto:10eroen.kuenen@tno.nl)).

We will first need to convert the format before using it.

We will use version “v5.1” of this dataset. If a new version becomes available during the project, we will stay with this version.

The provider keeps old versions around so the same reference data will be available to reproduce our results.

We will use the dataset as follows: Anthropogenic emissions for historical simulations of organic aerosols using a chemical transport model.

- **Global Fire Assimilation System daily biomass burning inventory v2.1**

(<https://eccad3.sedoo.fr/metadata/609>)

Owner of this dataset: ECMWF on behalf of European Commission; Contact: European Centre for Medium-range Weather Forecasts – Reading – United Kingdom – Miha Razinger ([miha.razinger@ecmwf.int](mailto:miha.razinger@ecmwf.int)).

The dataset can be used in the provided format without any conversion needed.

We will use version “v2.1” of this dataset. If a new version becomes available during the project, we will stay with this version.

The provider keeps old versions around so the same reference data will be available to reproduce our results.

We will use the dataset as follows: Fire emissions for historical simulations of organic aerosols using a chemical transport model.

- **EARLINET LIDAR** (<https://www.earlinet.org/index.php?id=125>)

Owner of this dataset: EARLINET LIDAR aerosol profile data are reported in the EARLINET Data base: <https://data.earlinet.org/> , and are accessible from this repository and from the ACTRIS Data Portal (<http://actris.nilu.no>). .

We will first need to convert the format before using it.

The provider keeps old versions around so the same reference data will be available to reproduce our results.

We will use the dataset as follows: To compare vertical and vertically integrated aerosol optical properties with model outputs.

- **AOD (550 nm) which combines several satellite products.** ([https://nsdc.fmi.fi/data/data\\_aod](https://nsdc.fmi.fi/data/data_aod))

Owner of this dataset: L. Sogacheva et al.: Merging regional and global aerosol optical depth records from major available satellite products, Atmos. Chem. Phys., 20, 2031–2056, <https://doi.org/10.5194/acp-20-2031-2020>, 2020. .

The dataset can be used in the provided format without any conversion needed.

The provider keeps old versions around so the same reference data will be available to reproduce our results.

We will use the dataset as follows: To compare global AOD at 550 nm with model outputs.

- **CAMS global reanalysis (EAC4)** (<https://ads.atmosphere.copernicus.eu/>)

Owner of this dataset: ECMWF, Copernicus Atmosphere Monitoring Service (CAMS), Inness et al. (2019), <http://www.atmos-chem-phys.net/19/3515/2019/>.

The dataset can be used in the provided format without any conversion needed.

We will use version “EAC4” of this dataset. If a new version becomes available during the project, we will stay with the old version.

The provider keeps old versions around so the same reference data will be available to reproduce our results.

We will use the dataset as follows: Build boundary conditions for regional atmospheric model simulations in WP4.

The following non-reference datasets will be considered for re-use:

- **GHOST (Globally Harmonised Observational Surface Treatment) database** (no persistent identifier available yet (GHOST will be published during 2023))

Owner of this dataset: Contact: Barcelona Supercomputing Center (dene.bowdalo@bsc.es). The owners of the dataset will collaborate on this project.

The dataset can be used in the provided format without any conversion needed.

We will download the data once it becomes available in 2023.

It is a fixed dataset; changes will not influence reproducibility of our results.

Only part of the dataset will be used; any filtering or selection will be well documented.

We will use the dataset as follows: GHOST database gathers observations from the following networks: ACTRIS, AERONET, AMAP, CAMP, Canada NAPS, CAPMoN, Chile SINCA, EANET, AirBase, EEA\_AQ\_eReporting, EMEP, EUCAARI, EUSAAR, HELCOM, HTAP, independent dataset from EBAS, Japan NIES, Mexico CDMx, NILU, NOAA ESRL, OECD, SEARCH, UK Air, UK DECC, US EPA AQS, US EPA CASTNET, US NADP AMNet, US NADP AmoN, WMO WDCA, WMO WDCGG, WMO WDCRG. The data will be used for model evaluation.

### 3.4 Project generated datasets

The FOCI project will generate the following datasets as part of its deliverables:

- Harmonized dataset of in-situ and column integrated aerosols and gas-phase species (D1.1)
- Database of advanced aerosol particles optical properties (D1.2)
- In-situ long-term datasets of aerosol, BVOC, CO2 fluxes and meteorological data (D2.1)
- Evaluation of specific forcings impact in ESMs (D3.4)
- Sensitivity tests and evaluation data outputs from regional simulations (D4.2, D4.3)
- New global emission scenarios for anthropogenic sources for GHGs including non-CO2 forcings (D5.1)
- New regional emission scenarios for anthropogenic sources for GHGs including non-CO2 forcings (D5.2)

- Refined emissions inventory of BVOC emissions for climate projections with coupled models (D5.3)
- New urban greening emission scenario for biogenic emissions (D5.4)
- High resolution modelling projections for near and mid future (D6.1)
- Datasets on future climate extremes concentrations for mitigation options assessment (D6.4)

There will be a host of ancillary datasets associated with the above. These include input emissions and meteorological (including terrain and land cover) datasets used for regional and global models.

As part of the project's data management, a questionnaire will be sent out to all project partners to gather more information about the data's metadata, format, size, and access. This will be checked against the main principles set out in Section 3 before being summarized in future versions of this document.

## **4. DATASETS TO SUPPORT STAKEHOLDERS**

### **4.1 Datasets to support policy and decision makers**

To ensure the usefulness of datasets for stakeholders, Task 8.4 (Outreach and engagement with the policy community) will in close collaboration with the Climate and Clean Air Coalition (CCAC). For example, policy relevant scenarios will be developed in WP5 and WP7 which will be communicated through the CCAC working groups to the community of CCAC partners focussing on national scale action and planning. Teams involved in WP7 and WP8 (e.g., SEI and TAU) are members of the CCAC Science Advisory Panel and are involved in implementing the national planning activities of the CCAC. Datasets generated by the project will feed into country specific policy recommendations, scientific assessments, updates, and reports and will be communicated to supporting national action and planning initiatives (SNAP). Where relevant synthesis derived from the analysis of datasets will be support other policy relevant mechanisms (e.g., communication to the NDC partnership, joint communications on policy options with Copernicus, presentations at UNFCCC meetings).

### **4.2 Datasets to support of operational services**

Task 8.5 (Use of results in support of operational services) will address the use of project datasets and other outputs and support operational sector-oriented services (e.g., climate services, operational forecasts, energy, health and urban services). The network of participating integrational organizations and NGOs will be widely used for the user mapping and deliverables verification including key datasets. The project will be guided by structured stakeholder consultations with the health and energy community to determine the critical gaps in existing datasets and information required in support of health assessments and energy sector transition. For example, future scenario runs of air pollution and climate will be shaped by the requirements of the joint WMO-WHO Study Group on Integrated Health Services and the WMO Study Group on Integrated Energy Services.

These datasets will feed into the evaluation of health risks associated with the climate and air pollution (global burden of disease) and on the impact assessment of FOCI scenarios on energy supply and demand. This project specifically provides a unique opportunity for close cooperation between regional and global modelling communities to generate multiscale datasets relevant for air quality and climate. In particular, we will work closely with the operational ESM community also through engagement of the modelling centres outside of Europe (USA, Canada) building on WMO coordinating groups (e.g., Standing Committee on Earth System Modelling). In fact, WMO Study Groups on Integrated Energy Services and Integrated Urban services will feed into defining scenario runs which will generate new high-resolution climate and atmospheric composition datasets. Engagement with the operational weather forecasting and climate communities will also ensure the legacy of new datasets resulting from the project.

### **4.3 Intellectual Property Rights and data protection**

In general, intellectual property rights (IPR) are pursued for new methods and data products. IPR Owner(s) are the consortium partner(s) that will own the result. To ensure that all new knowledge and intellectual property generated by the project is managed correctly and adequately protected, the consortium has agreed upon a Consortium Agreement, including ownership and pre-existing rights. Only after such existing and new IPR has been protected and upon achieving the generic and broadly applicable results as pre-agreed by the consortium, those results that are non-confidential will be made public.

### **4.4 Quality assurance and control management of datasets**

Task 8.3 which manages the DMP, will work closely with Tasks 9.3 (D9.1 QA/QC and Risk Management) to ensure agreed scientifically acceptable criteria and protocols are applied to the datasets that arise from the project. Specifically, this will be important for datasets related to measurements, modelling and integration of methods. Consequently, these procedures will ensure the highest quality assurance and control (QA/QC) of all datasets which will feed into project deliverables, publications, and other results and outputs including designing an effective risk management strategy.

This Data Management Plan will also complement deliverable D8.2 Communication, dissemination, and exploitation plan as together with Task 9.3 (D9.1), they form a strong basis for ensuring successful, timely and high-quality implementation of all project activities.

As part of the QA/QC process for datasets, there will be periodical updates of the DMP (with a frequency of at least six monthly) incorporating new details as they emerge during the implementation of the project, in consultation with all relevant partners.

## 5. FAIR DATA

This section outlines the main idea and guidelines for ensuring the FOCI datasets adhere to the Findable, Accessible, Interoperable and Reusable (FAIR) principles.

### 5.1 Making data findable

Following the FAIR principles, metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the FAIRification process. This includes the use of a globally unique and persistent identifier that is registered or indexed in a searchable resource (e.g., DOI) and the use of rich metadata describing the dataset itself.

The data produced by the FOCI project (therefore, not the re-used data) will initially be stored on data servers by individual project partners and be accessible by request only. However, there is the intention to publish the data sets that are mature and relevant for the wider science community. In these cases, the relevant data and metadata will be archived on external repositories and reference to these datasets will be made with DOI numbers, as much as possible. This will ensure that subsequent work done using these datasets can be conducted by downloading the data from data repositories without asking individual permissions from authors and they can be referenced with DOI links. The FOCI website will also provide references to all the deliverables (reports, data sets, and others), once they have been approved by the Commission.

Further details for the specific datasets will be included in future versions of this DMP.

### 5.2 Making data accessible

Once a user finds the required data, they need to know how the data can be accessed, possibly including authentication and authorization. This means that (meta)data are retrievable by their identifier using a standardized communications protocol (e.g., ftp or a web-based data portal). The protocol must be open, free, and universally implementable and allow for an authentication and authorization procedure, where necessary.

FOCI will produce very large data sets on the order of terabytes of data. Most of the data will be for internal usage and project partners have sufficient internal storage and network bandwidth to deal with these data sets and to exchange data between project partners. Data storage of 160 TB in NAS (Network Attached Storage) have been acquired and installed at CU for the exchange of data, data sharing and coordinated postprocessing and analysis.

For the data sets that can be made available outside the project, specific solutions will be put in place to ensure the data can be shared with the wider science community. The implementation of this will be documented in future versions of this DMP.

### **5.3 Making data operable**

It is common practice that data is combined or integrated with other data by the users. This also means that the data need to interoperate with existing applications or workflows for analysis, storage, and processing. Therefore, (meta)data shall use a formal, accessible, shared, and broadly applicable language for knowledge representation and use vocabularies that follow FAIR principles.

Some of the modelling systems used in FOCI are already part of various international model intercomparison frameworks (e.g., CMIP) and as such, are already used to interoperable requirements for model inputs and outputs. For other modelling systems care shall be taken to adhere to these FAIR principles.

### **5.4 Increase data re-use**

The aim of FAIR is to optimize the re-use of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings. This means that (meta)data are richly described with a plurality of accurate and relevant attributes, are released with a clear and accessible data usage license, are associated with detailed provenance, and meet domain-relevant community standards.

As already stated above, relevant datasets produced by FOCI will be made public, either as deliverable or as part of a data describing article in the peer-reviewed literature. For smaller datasets, platforms like Zenodo (<https://zenodo.org/>) can be used. For larger datasets, details will follow in later versions of this DMP.

## **6. ALLOCATION OF RESOURCES**

The costs required for making the data collected/generated FAIR have been included in the budget of the project. While Charles University (CU), as project coordinator, have the overall responsibility for the delivery of the project, the European Centre for Medium-Range Weather Forecasts (ECMWF) and the Task 8.3 team have the primary responsibility for creating, updating and delivering the data management plan. However, all project partners are involved, either in support of the overall data management plan or by ensuring their own contributions adhere to the FAIR principles. The related costs have therefore been included in the overall contribution of each partner.

## **7. CONCLUSIONS**

This first version of this Data Management Plan sets out the principles for a proper management of the various datasets used and produced in the project as well as a first overview of especially the re-used data sets. The next step will consist in collecting more details from FOCI's partners about the datasets they are, or will be, producing. The provided information will be checked against the FAIR principles and adjustments will be made,



where needed, before publishing the results. There will be regular updates of the DMP with a frequency of at least six months, and it is expected that a next version of the DMP will appear at M12 of the project.

## 8. REFERENCES

- [1] Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020: [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)
- [2] Horizon 2020 Online Manual: [https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination\\_en.htm](https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm)
- [3] Horizon Europe Programme Guide: [https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide\\_horizon\\_en.pdf](https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf)
- [4] Horizon Europe Model Grant Agreement
- [5] Horizon Europe Model Consortium Agreement

## APPENDICES

### Appendix 1. List of Partners

No	Role	Short name	Legal name	Country
1	COO	CU	UNIVERZITA KARLOVA	CZ
2	BEN	MPI-C	MAX-PLANCK-GESELLSCHAFT ZUR FORDERUNG DER WISSENSCHAFTEN EV	DE
3	BEN	FMI	ILMATIETEEN LAITOS	FI
4	BEN	UHam	UNIVERSITAET HAMBURG	DE
5	BEN	CSIC	AGENCIA ESTATAL CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS	ES
6	BEN	KNMI	KONINKLIJK NEDERLANDS METEOROLOGISCH INSTITUUT-KNMI	NL
7	BEN	BSC	BARCELONA SUPERCOMPUTING CENTER-CENTRO NACIONAL DE SUPERCOMPUTACION	ES
8	BEN (IO)	WMO	WORLD METEOROLOGICAL ORGANIZATION	CH
9	BEN	UHeI	HELSINGIN YLIOPISTO	FI
10	BEN	TAU	TEL AVIV UNIVERSITY	IL
11	BEN (IO)	ECMWF	EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS	UK
12	BEN (IO)	WHO	WORLD HEALTH ORGANIZATION	CH

13	BEN	ARIANET	ARIANET SRL	IT
14	BEN	SU	STOCKHOLMS UNIVERSITET	SE
15	AP	UH	THE UNIVERSITY OF HERTFORDSHIRE HIGHER EDUCATION CORPORATION	UK
16	AP	SEI	UNIVERSITY OF YORK	UK
17	AP	WEMC	WORLD ENERGY & METEOROLOGY COUNCIL	UK

## Appendix 2. List of Abbreviations

ACTRIS	Aerosol, Clouds and Trace Gases Research Infrastructure
AERONET	Aerosol Robotic Network
BVOC	Biogenic Volatile Organic Compound
C3S	Copernicus Climate Change Service
CA	Consortium Agreement
CAMS	Copernicus Atmosphere Monitoring Service
CCAC	Climate and Clean Air Coalition
CDE	Communication, Dissemination and Exploitation
CMIP	Coupled Model Intercomparison Project
D (number)	Deliverable (number)
DMP	Data Management Plan
DOI	Digital Object Identifier
EARLINET	European Aerosol Research Lidar Network
EC	European Commission
EMEP	European Monitoring and Evaluation Programme
ESM	<i>Earth System Model</i>
FAIR	Findable, Accessible, Interoperable and Reusable
FOCI	non-CO2 Forcers and their Climate, Weather, Air Quality and Health Impacts
GA	Grant Agreement
GHG	Greenhouse Gas
IPCC	Intergovernmental Panel on Climate Change
IPR	Intellectual Property Rights
IMPROVE	Interagency Monitoring of Protected Visual Environments
PCG	Project Coordination Group
PMC	Project Management Committee
PO	Project Office
QA/QC	Quality Assurance/Quality Control
SNAP	Supporting National Action and Planning
UKRI	UH Research and Innovation
UNFCCC	United Nations Framework Convention on Climate Change
WMO/GAW	World Meteorological Organization Global Atmospheric Watch programme
WP	Work Package

