



Deliverable 8.3 Data Management Plan

Authors: Sander Houweling (VUA)
Date: 17/6/2025
Dissemination: Public
Work package: 8.1
Version: 2.0



Funded by
the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101183460

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

Investigating Methane for Climate Action (IM⁴CA)

Call: HORIZON-CL5-2024-D1-01

Topic: HORIZON-CL5-2022-D1-01-01

Type of action: HORIZON Research and Innovation Actions

Granting authority: European Climate, Infrastructure and Environment Executive Agency

Project starting date: 01/01/2025

Project end date: 31/12/2028

Project duration: 48 months

Contact: Prof. dr. ir. Sander Houweling, Coordinator
Vrije Universiteit Amsterdam, The Netherlands

Document history:

Version	Author(s)	Date	Changes
1	S. Houweling (VUA)	20/4/2025	First draft version
2	S. Houweling (VUA)	17/6/2025	Internal review corrections

Internal review: Alex Vermeulen (ICOS-ERIC), Margareta Hellström (ULUND)

Table of Contents

1	PLAN OVERVIEW	5
2	DATA DESCRIPTION.....	6
2.1	WILL YOU COLLECT AND/OR PROCESS PERSONAL DATA IN THIS PROJECT?.....	6
2.2	WILL YOU USE EXISTING DATA? IF YES, WHAT IS THEIR SOURCE?.....	6
2.3	WILL YOU COLLECT OR PRODUCE NEW DATA? IF YES, PLEASE DESCRIBE HOW.	6
2.4	WHAT KINDS OF OUTPUT DATASETS WILL YOU PRODUCE IN THIS PROJECT?.....	7
2.5	HOW MUCH DIGITAL DATA STORAGE WILL YOUR PROJECT REQUIRE?	7
2.6	WILL YOU TAKE MEASURES TO ENSURE DATA QUALITY?	8
3	LEGAL AND ETHICAL REQUIREMENTS, CODES OF CONDUCT	8
3.1	WHAT LEGISLATION APPLIES TO YOUR RESEARCH PROJECT? PLEASE TICK THE RELEVANT BOXES FOR YOUR PROJECT	8
3.2	DO YOU REQUIRE APPROVAL OF AN ETHICAL COMMITTEE FOR THIS PROJECT?.....	8
3.3	WILL YOU WORK WITH DATA FOR WHICH INTELLECTUAL PROPERTY AND/ OR CONFIDENTIALITY ARE AN ISSUE?.....	8
3.4	DO YOU PLAN ON GENERATING A MARKETABLE PRODUCT FROM YOUR RESEARCH PROJECT?.....	8
4	STORAGE AND BACK-UP DURING THE RESEARCH PROCESS	9
4.1	WHAT MEASURES WILL YOU TAKE TO SECURE AND PROTECT DATA DURING THE RESEARCH PROCESS?9	
4.2	IS IT NECESSARY TO TRANSFER THE (PHYSICAL OR DIGITAL) DATA ASSETS TO OTHER LOCATIONS OR RESEARCH PARTNERS? IF YES, HOW IS THE FILE TRANSFER SECURED.	9
5	DATA ARCHIVING AND PUBLISHING	9
5.1	WHICH DATA ASSETS WILL BE ARCHIVED AND WHICH WILL BE PUBLISHED?.....	9
5.2	WHERE WILL YOU ARCHIVE YOUR DATA ASSETS?	10
5.3	FOR HOW LONG WILL THE DATA BE AVAILABLE IN THE ARCHIVE?	10
5.4	WHERE WILL YOU PUBLISH YOUR DATA ASSETS?	10
5.5	HOW WILL YOU ENSURE YOUR DATASET GETS A PERSISTENT IDENTIFIER?.....	10
5.6	WILL YOU REGISTER YOUR DATASETS IN AN ONLINE REGISTRY OTHER THAN PURE? IF YES, WHERE? 10	
5.7	ARE THERE RESTRICTIONS TO DATA PUBLISHING?.....	10
5.8	WHEN WILL YOU SHARE THE DATA?	11
5.9	PLEASE INDICATE THE LICENSE AND/ OR TERMS OF USE UNDER WHICH YOU SHARE YOUR DATA.	11
6	DOCUMENTATION	11
6.1	WHAT DOCUMENTATION WILL ACCOMPANY THE DATA?.....	11
6.2	WHAT METADATA WILL ACCOMPANY THE DATA?	11
6.3	WILL SEARCH KEYWORDS BE PROVIDED IN THE METADATA TO OPTIMIZE DISCOVERY AND POTENTIAL RE-USE? 11	
6.4	WILL SEARCH KEYWORDS BE PROVIDED IN THE METADATA TO OPTIMIZE DISCOVERY AND POTENTIAL RE-USE? 11	
6.5	WHAT METHODS, SOFTWARE OR HARDWARE ARE NEEDED TO ACCESS AND USE YOUR DATA?.....	12
7	OTHER RESEARCH OUTPUTS	12
7.1	HOW ARE OTHER RESEARCH OUTPUTS SUCH AS SOFTWARE MANAGED?.....	12
7.2	HOW IS THE FAIR SHARING OF SOFTWARE MANAGED AND GUARANTEED?	12
8	DATA MANAGEMENT RESPONSIBILITIES AND RESOURCES.....	12
8.1	WHO WILL BE RESPONSIBLE FOR MANAGEMENT OF THE DATA ASSETS DURING THE PROJECT? PLEASE SPECIFY THEIR NAME, POSITION, ROLE IN THE PROJECT, AND FACULTY/ INSTITUTION/ GROUP.....	12
8.2	WHO WILL BE RESPONSIBLE FOR MANAGEMENT OF THE DATA ASSETS AFTER COMPLETION OF THE PROJECT (E.G. THE PROJECT LEAD/ DEDICATED DATA MANAGER/ DEPARTMENT HEAD)?	13

8.3 FOR DATA THAT ARE ONLY AVAILABLE UPON REQUEST, WHAT METHODS WILL BE USED TO HANDLE REQUESTS FOR ACCESS AND HOW WILL DATA BE MADE AVAILABLE TO THOSE REQUESTING ACCESS?13

8.4 WHAT RESOURCES (FOR EXAMPLE FINANCIAL AND TIME) WILL BE DEDICATED TO RESEARCH DATA MANAGEMENT?13

1 Plan Overview

A Data Management Plan created using DMPonline

Title: Investigating Methane for Climate Action (IM4CA)

Creator: Sander Houweling

Affiliation: Vrije Universiteit Amsterdam

Principal Investigators:

Sander Houweling (Coordinator, UU, The Netherlands)

Thomas Röckmann (Co-coordinator, VU, The Netherlands)

Mirosław Zimnoch (AGH, Poland)

Mashesh Kumar Sha (BIRA, Belgium)

Jean-Daniel Paris (CEA, France)

Jean Sciare (CYI, Cyprus)

Richard Engelen (ECMWF, Germany)

Leonidas Paroussos (E3-Modelling AE, Greece)

Tuula Aalto (FMI, Finland)

Torsten Sachs (GFZ, Germany)

Lena Hoglund-Isaksson (IIASA, Austria)

Sebastian Iancu (INCAS, Romania)

Alexandru Dandocsi (INOE, Romania)

Sönke Zaehle (MPI-GBC, Germany)

Rona Thompson (NILU, Norway)

Hannah Lee (NTNU, Norway)

Eric Hamanou (Science Partners, France)

Ilse Aben (SRON, The Netherlands)

Hugo Denier van der Gon (TNO, The Netherlands)

Hartmut Bösch (Uni. Bremen, Germany)

Manuel Gloor (Uni. Leeds, United Kingdom)

Marko Scholze (Uni. Lund, Sweden)

Andreas Stohl (Uni. Vienna, Austria)

Marielle Sauniois (UVSQ, France)

Funder: European Commission

Template: 1 - VU DMP template 2021 v1.4

Project abstract:

The IM⁴CA proposal unites leading European methane experts in a concerted effort to establish the scientific fundament needed to bring the climate forcing of methane under control. Specific objectives are to: 1) Strengthen methane mitigation policy world-wide with actionable information on local methane emissions and key driving processes, 2) Provide the EU with the measurement and modeling capacity needed to monitor its methane emissions and assess its progress towards the 30% emission reduction target of the European methane strategy and the global methane pledge, 3) Explore and understand climate feedbacks on natural methane sources and sinks, and 4) Improve the accuracy of climate scenarios by resolving the controversy about the causes for the recent growth rate variations in global methane. To achieve these objectives we address a selection of key uncertainties that have

thus far limited the progress towards these goals by: i) building up critical new infrastructure for monitoring methane emissions in Europe and Tropical Africa, ii) developing methodology for efficient use of existing and upcoming satellites for measuring methane and the land surface properties needed for characterizing and attributing its emissions, and iii) translating the knowledge obtained into reliable projections of future methane and efficient emission mitigation scenarios. In doing so, the IM4CA project will enable a breakthrough in meeting the work program challenge of enhancing the quantification and understanding of natural and anthropogenic methane emissions and sinks. It will provide enhanced European assessment capacity of short- and long-term changes in methane sources and sinks integrating information from multi platforms and novel observations and transfer that capacity into actionable information needed to combat climate change.

ID: 178584

Start date: 01-01-2025

End date: 01-01-2029

Grant number / URL: 101183460

2 Data description

2.1 Will you collect and/or process personal data in this project?

The use of personal data will be limited to pictures and contact information for use on the website, contact lists, eu-portal, meeting minutes, conference abstracts, co-author lists, etc. We minimize the information that is collected and treat this information in full compliance with GDPR regulations.

2.2 Will you use existing data? If yes, what is their source?

We will use existing spatio-temporally explicit data on the surface fluxes of methane, provided by emission inventories and models. These data will be supported by data of many other auxiliary variables (land surface characteristics, meteorological variables, etc.) that influence the emissions and can be used to estimate them. In addition, we will use existing surface and satellite measurements of methane and other supporting tracer gases. These other trace gases are important, in particular for estimating methane sinks in the atmosphere.

Most datasets can be obtained freely from public data archives, either as part of scientific publications, or provided by organizations (space agencies, the integrated carbon observing system, the world data center of greenhouse gases, etc.). Some existing data are owned by consortium members and will be used by other project partners.

2.3 Will you collect or produce new data? If yes, please describe how.

Various activities in the project will generate new data. These concern either new measurements or new model output datasets.

The measurements will be made using instruments that are in part purchased with project funding, or in use by partners already prior to the start of the project. Measurements will be collected from instruments operating in situ, on ground, or from drones and aircrafts. In addition, flasks are

sampled in situ and transported to laboratories of project partners for analysis. New satellite data will be collected by requesting the operators to point the instrument to specific measurement targets on ground.

Model datasets of various parameters will be generated by emission inventories, process-based models, atmospheric transport (and chemistry) models, inverse models, and integrated assessment models.

2.4 What kinds of output datasets will you produce in this project?

The project will produce primary data, for example on methane sources and sinks specified by time, location, and sector including uncertainties. These data will be synthesized into secondary data products, such as regionally and temporally integrated emission totals that are used in our deliverable reports.

The data will mostly be stored in Netcdf format, preferably following the cf convention, together with appropriate meta data to describe the dataset. Some datasets might become available in text format, for example as CSV files. Model data are usually gridded in the space and time dimensions of the model domain. Measurements are time series at specified horizontal and vertical coordinates where the atmospheric samples are taken.

2.5 How much digital data storage will your project require?

The estimated total digital storage during the active project stage is up to 5 TB. The storage requirements will be distributed over different supercomputers used within the consortium. Only a fraction of these data (within 1 TB) will have to be archived as project output on publicly accessible data portals.

The data storage requirements will vary by data type, to which the following (rough) estimates apply:

Measurements:

- in-situ: a few GB
- satellite: ~100 GB per month of data in the case of TROPOMI. These primary data do not need to be stored but can be processed into smaller dataset for use in models (100 GB for global inversions spanning several years)

Model input datasets:

- 10 GB for surface fluxes of a single tracer from a global model run spanning several decades
- 40 GB of meteorological driving fields for a global model run for a single year

Model output datasets:

- 10 GB for surface fluxes of a single tracer from a global model run spanning several decades
- 4 GB for mixing ratio fields for a global model run for a single year

2.6 Will you take measures to ensure data quality?

Measurement and model data will be made available as much as possible with estimates of their uncertainties. Measurements will be calibrated using internationally agreed reference standards, and undergo quality control using automated and manual data filtering and flagging methods. Model results will be validated using the available measurements where possible. Where possible, models be compared to each other to check consistency.

3 Legal and ethical requirements, codes of conduct

3.1 What legislation applies to your research project? Please tick the relevant boxes for your project.

- Other, please specify below

We will not use or generate sensitive data that require any specific legislation.

3.2 Do you require approval of an ethical committee for this project?

There is no ethical restriction on the collection of air samples anywhere in the world. Therefore, no approval of an ethical committee is required.

3.3 Will you work with data for which intellectual property and/ or confidentiality are an issue?

No, this is not an issue. Some intellectual property rights have been defined in the consortium agreement. However, they refer to the use of models, rather than the data they generate.

3.4 Do you plan on generating a marketable product from your research project?

No, this is not planned.

4 Storage and back-up during the research process

4.1 What measures will you take to secure and protect data during the research process?

The most important locations where IM4CA data will be archived are the ICOS Carbon Portal and the Surf research drive.

Surf research drive is backed up daily and back ups are kept for 30 days. This includes disaster recovery backup to protect against hardware failure, with a recovery point objective of 24 hours. At ICOS Carbon Portal all data is securely copied using the EUDAT CDI B2SAFE service, that stores two live copies of each digital artefact at two locations (one at CP, Sweden, and one at IT4I, Czech Republic). Next to that ICOS Carbon Portal stores its own copy of the data, with a separated backup at two different buildings at Lund University, Sweden.

On Surf Research Drive all project partners can gain access after registration. The login is protected using two-factor authentication. At the ICOS Carbon Portal only those project partners that are registered as such get access to the tools and services related to archiving and publishing IM⁴CA datasets and related metadata.

4.2 Is it necessary to transfer the (physical or digital) data assets to other locations or research partners? If yes, how is the file transfer secured.

Yes, data transfer is important. Research drive provides a platform for exchanging datasets during the execution of the project, as long as they are not in their final published form yet. Final and published datasets are made available through the ICOS Carbon Portal. Data integrity is ensured by using the SHA256 checksums used at all transfers by sender and receiver and inclusion of this information in the object Handle PID. Data objects can be versioned, joined into collections with unique Handle PIDs and minted a Datacite DOI.

5 Data archiving and publishing

5.1 Which data assets will be archived and which will be published?

All data will be documented in deliverable reports. All those reports will be publicly accessible through the project website. Links to datasets that are stored on the ICOS Carbon Portal will appear on the website also. The intention is to publish most model simulations that we want to archive. In this process the data volume will have to be reduced to retain only the most important data that are expected to be re-used. All measurements will be archived regardless of whether they are published or not. The intention is to publish as many as possible, for example as part of model assessments that make use of the data. The remaining data, which 'do not fit in' our publications will be archived without publication, but with information in a corresponding deliverable report describing the dataset and how it can be obtained.

5.2 Where will you archive your data assets?

Final datasets will be published using FAIR principles and archived on the ICOS Carbon Portal.

5.3 For how long will the data be available in the archive?

ICOS is a long-term research infrastructure, legally presented by ICOS ERIC, that is foreseen to exist permanently, for at least 15 years. ICOS Carbon Portal will preserve data and metadata for the lifetime of ICOS and beyond. After the termination of ICOS all data and metadata will be transferred to a static website and/or other trusted repositories during the decommissioning period of at least one year.

5.4 Where will you publish your data assets?

Data assets will be published in peer reviewed scientific publications as much as possible. References to these papers will appear on the project website. These and other datasets are documented also in project deliverables that are made available publicly through the project website. Links to datasets will also be made available from the website. Besides the ICOS Carbon Portal, published datasets may also be archived in parallel at other repositories (Pangea, Zenodo, or local RDM archives such as Yoda at VU)

5.5 How will you ensure your dataset gets a persistent identifier?

All data to be retained will be provisioned with a persistent identifier, in the default case of using the ICOS repository this will at minimum be a Handle PID (ePIC), and in most cases also a Datacite DOI PID will be minted. In some cases data will be published using other public archives with similar functionality that mints PIDs or DOIs, like Pangea or Zenodo. In case Pangea or Zenodo are used, their data management practices apply, that we assume are already accepted as best practice.

5.6 Will you register your datasets in an online registry other than PURE? If yes, where?

Datasets will be registered and published through the ICOS Carbon Portal. ICOS Carbon Portal is indexed by major portals of portals like GEOSS and Google Dataset Search. Datasets will also be findable on the project website and other data repositories in case they are used (Pangea, Zenodo, Yoda).

5.7 Are there restrictions to data publishing?

There are no restrictions to the publication of data, with the exception of data that await the acceptance of a scientific publication in which they are described / used.

5.8 When will you share the data?

Data will be published immediately after scientific publication. Data that have not been published during the project will become accessible after completion of the project.

5.9 Please indicate the license and/ or terms of use under which you share your data.

All raw and higher level data products provided will follow the ICOS data licence which is CC-BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>)

6 Documentation

6.1 What documentation will accompany the data?

Dataset are documented in corresponding deliverable reports and meta data that is made available with the data. On the ICOS Carbon Portal the associated metadata linked to the data is stored in a RDF database, based on the OWL open ontology, that is part of the Carbon Portal RDF database.

6.2 What metadata will accompany the data?

The data will be foreseen with detailed provenance data on instrumentation and methods used and includes detailed information on location, variables, units, data providers and their roles, all modelled over time. Keywords are chosen that are commonly used and unambiguous in the research field.

6.3 Will search keywords be provided in the metadata to optimize discovery and potential re-use?

Yes, keywords will be provided, users are encouraged to select keywords from the GCMD (Global Change Master Directory) standard. In the Carbon Portal data portal files can be selected based on a faceted search that includes the keywords, data type and other metadata elements.

6.4 Will search keywords be provided in the metadata to optimize discovery and potential re-use?

Yes, at the data portals used like Zenodo, Pangaea and ICOS Carbon Portal all metadata is accessible to both human and machine to machine communication, at the Carbon Portal using the following standards: iso19115, Schema.org, and by content negotiation as xml, json-ld, rdf/turtle, rdf/xml. All metadata is served from the landing pages and the open sparql endpoint.

6.5 What methods, software or hardware are needed to access and use your data?

The most common use of data makes use of programming languages (Python, Fortran, C++) for efficient processing of large volumes of data. Measurement data that are provided in CSV format can be read also by software tools such as Microsoft Excel and libraries such as Python Pandas. Documentation of the data will be made available in pdf format, readable by any pdf viewer.

7 Other research outputs

7.1 How are other research outputs such as software managed?

Software developed in the project will in principle be provided as open source through platforms such as Github using open licenses such as GPL. Exceptions are existing model codes that are background IP and modifications to that software. Workflows and protocols developed in the project will be published as project results through the deliverables and will be open without restriction.

7.2 How is the FAIR sharing of software managed and guaranteed?

IM⁴CA will cooperate with Horizon-Europe project EVERSE on the FAIR sharing of software and adopt their recommendations on software management, whenever considered appropriate and executable.

8 Data management responsibilities and resources

8.1 Who will be responsible for management of the data assets during the project? Please specify their name, position, role in the project, and faculty/ institution/ group.

All project partners are responsible for the data they generate in the project. Work package leads are responsible for the storage of final datasets that are made available from the work performed in their work package on the ICOS Carbon Portal. The coordinator provides information on the data storage platform that are used and maintains the research drive environment. Partner ULUND hosts the ICOS Carbon Portal and assists in the storage of final datasets there and the tools available on the Carbon Portal for interested users to find and explore the data.

Work package leads:

WP1: Thomas Röckmann (UU, The Netherlands), Jean-Daniel Paris (CEA, France)

WP2: Torsten Sachs (GFZ, Germany), Tuula Aalto (FMI, Finland)

WP3: Manuel Gloor (Uni. Leeds, UK), Hartmut Bösch (Uni. Bremen, Germany)

WP4: Andreas Stohl (Uni. Vienna, Austria), Hugo Denier van der Gon (TNO, The Netherlands)

WP5: Marielle Saunois (UVSQ, France), Rona Thompson (NILU, Norway)

WP6: Lena Höglund-Isaksson (IIASI, Austria), Sergey Gromov (MPI-C, Germany)

WP7: Ernest Koffi (ECMWF, Germany), Roxana Petrescu (VU, The Netherlands)

WP8: Sander Houweling (VU, The Netherlands), Thomas Röckmann (UU, The Netherlands)

Coordination team:

Sander Houweling (VU, The Netherlands) [Data manager at Surf Research Drive, and the web portal]

Mekky Zaidi (VU, The Netherlands)

Contacts at ICOS Carbon Portal:

Alex Vermeulen (ICOS ERIC, Sweden) [Data manager at ICOS Carbon Portal]

Margareta Hellström (ULUND, Sweden) [data steward at ICOS CP]

Ute Karstens (ULUND, Sweden) [data and modelling scientist]

8.2 Who will be responsible for management of the data assets after completion of the project (e.g. the project lead/ dedicated data manager/ department head)?

The project coordinator will be responsible for the data assets after completion of the project, supported by ICOS Carbon Portal staff for any maintenance work needed to ensure findability and accessibility of the data.

8.3 For data that are only available upon request, what methods will be used to handle requests for access and how will data be made available to those requesting access?

This will only happen for data that are not yet published or are in the process of becoming available. Those data can be requested from the 'contact' facility on the website, or in direct contact with the responsible PI.

8.4 What resources (for example financial and time) will be dedicated to research data management?

ICOS Carbon Portal has an annual budget of about 2 M€ that allows the appointment of 15 fte for continuous development of the data services (5 fte), provision of Jupyter VREs (3 fte), data stewards (1 fte), data scientists (3 fte) and the costs for the Carbon Portal cloud services and B2SAFE. Budget is decided by the ICOS General Assembly, consisting of representatives of the currently 16 European member states. The IM4CA project supports partner ULUND with 12 PM of staff funding for making IM4CA data findable on the ICOS Carbon Portal and supporting data use.