Plan Overview

A Data Management Plan created using DMPonline

Title: Moisture Tracking for Climate Models: a Community Effort

Creator: Ruud van der Ent

Principal Investigator: Ruud van der Ent

Data Manager: Ruud van der Ent, Imme Benedict, Peter Kalverla

Project Administrator: Ruud van der Ent

Contributor: Imme Benedict, Chris Weijenborg, Peter Kalverla, Arie Staal

Affiliation: Delft University of Technology

Funder: Netherlands Organisation for Scientific Research (NWO)

Template: Data Management Plan NWO (September 2020)

ORCID iD: 0000-0001-5450-4333

Project abstract:

In this project a research community works together to further develop existing tools such that they will be capable to track fictitious water particles through the atmosphere within climate models. They aim to gain more insight into the pros and cons of different tools in terms of accuracy, computational speeds, user-friendliness and data storage. By practising Open Science-principles they will be able to use each other’s tools more easily, learn from each other and carry out improvements. New users will also be able to make better-informed decisions on which tool will be most applicable for their specific research questions.

ID: 133218

Start date: 01-01-2024

End date: 31-12-2024

Last modified: 21-09-2023

Grant number / URL: OSF23.1.029

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Moisture Tracking for Climate Models: a Community Effort

General Information

Name applicant and project number

Ruud van der Ent
OSF23.1.029

Name of data management support staff consulted during the preparation of this plan and date of consultation.

My faculty data steward, Xinyan Fan, has reviewed this DMP on Monday 11 September

1. What data will be collected or produced, and what existing data will be re-used?

1.1 Will you re-use existing data for this research?

If yes: explain which existing data you will re-use and under which terms of use.

- Yes

Yes, we will likely re-use existing data from ongoing moisture tracking intercomparison projects. There are no constraints regarding re-use. This will be ERA5 reanalysis data and CMIP climate model data. All in netcdf files and distributed under license CC BY 4.0.

1.2 If new data will be produced: describe the data you expect your research will generate and the format and volumes to be collected or produced.

This research will primarily generate new spatial-temporal numerical data of moisture sources, which is based on climate model input data. Specific parts will be used as scientific output, but more relevant to this open science project is the idea of generating 'standard data', so new users can try to reproduce this data using the provided source code and documentation of the moisture tracking models.
Whenever possible, we will use file formats suitable for long-term preservation and re-use of research data. In our choices, we will adhere to the guidance provided by 4TU.ResearchData: https://data.4tu.nl/s/documents/Preferred_File_Formats_2023.pdf
Table below provides an overview of the types of data which will be collected and the associated file formats.

Type of data	Format
Numerical	NetCDF (.nc)
Metadata	.txt format
source code	depending on programming languages. Often .py or .yml

1.3. How much data storage will your project require in total?

- >1000 GB

The amount of data that needs persistent storage is likely to be less than 1000 GB, but during the research several TBs will be used.

2. What metadata and documentation will accompany the data?

2.1 Indicate what documentation will accompany the data.

All datasets will be accompanied by README files providing documentation necessary for data re-use data. Guidance provided by 4TU.ResearchData will be followed when preparing the README files. Moreover, we will provide followed modelling protocols and source code of the specific model runs wherever possible.

2.2 Indicate which metadata will be provided to help others identify and discover the data.

All data supporting publications will be made openly available through4TU.ResearchData. 4TU.ResearchData is a trusted and certified research data repository (it has a Data Seal of Approval certification). All datasets will be accompanied by rich metadata (adhering to DataCite metadata standard) to ensure that they are findable. In addition, to further aid their discoverability, keywords describing the datasets will be added. 4TU.ResearchData is also using schema.org metadata, meaning that all datasets are indexed in Google Dataset Search. Every dataset will be also assigned a Digital Object Identifier (DOI), to make them citable and persistently available.

3. How will data and metadata be stored and backed up during the research?

3.1 Describe where the data and metadata will be stored and backed up during the project.

- Institution networked research storage

For data sharing during the project we will set up a dedicated OneDrive for Business through TU Delft: https://storagefinder.tudelft.nl/package/6/
Since several institutions are involved we cannot fully control how each institution will handle data (temporary data possibly exceeding 1TB) and metadata, but we aim to keep things as central as possible on the dedicated OneDrive. All institutions involved are universities. We have provisional commitments of: TU Delft, WUR, Utrech University, University of Bergen, University of Santiago de Compostela, Vigo University, Stockholm University, Colorado State University, Ghent University and Karlsruhe Institute of Technology.
For the main software WAM2layers code will continue to be maintained on a dedicated GitHub version control system. In case WAM2layers needs temporary storage >1TB, this will be stored on local servers (Project Drive) maintained and automatically backed up by TU Delft ICT. Data can be recovered with the help of TU Delft ICT services in the event of an incident.

3.2 How will data security and protection of sensitive data be taken care of during the research?

- Not applicable (no sensitive data)

4. How will you handle issues regarding the processing of personal information and intellectual property rights and ownership?

4.1 Will you process and/or store personal data during your project?

If yes, how will compliance with legislation and (institutional) regulation on personal data be ensured?

- No

4.2 How will ownership of the data and intellectual property rights to the data be managed?

The datasets underlying the published papers will be publicly released following NWO’s policies. During the active phase of research, the lead applicant from TU Delft will oversee the access rights to data (and other outputs), as well as any requests for access from external parties. They will be released publicly no later than at the time of publication of corresponding research papers.

5. How and when will data be shared and preserved for the long term?

5.1 How will data be selected for long-term preservation?

- All data resulting from the project will be preserved for at least 10 years

All data supporting publications will be made openly available through4TU.ResearchData. 4TU.ResearchData is a trusted and certified research data repository (it has a Data Seal of Approval certification), and ensures that research data will be preserved for at least 15 years.

5.2 Are there any (legal, IP, privacy related, security related) reasons to restrict access to the data once made publicly available, to limit which data will be made publicly available, or to not make part of the data publicly available?

If yes, please explain.

- No

5.3 What data will be made available for re-use?

- All data resulting from the project will be made available

5.4 When will the data be available for re-use, and for how long will the data be available?

- Data available as soon as article is published

All research data underpinning research papers will be made publicly available by depositing at 4TU.ResearchData no later than at the time of the publication of the corresponding research article.

5.5 In which repository will the data be archived and made available for re-use, and under which license?

The datasets underlying the published papers will be published at 4TU.ResearchData, which is a trusted and certified research data repository (Data Seal of Approval certification). All datasets will be licensed under a CC-BY licence which requires attribution/credit for the original creation, while at the same time ensures broadest possible re-use. All datasets will be accompanied by rich and descriptive metadata, compliant with DataCite metadata schema, to ensure that all datasets are findable and accessible online. https://data.4tu.nl/info/en/

5.6 Describe your strategy for publishing the analysis software that will be generated in this project.

The developed software and codes presented in academic papers will be shared on GitHub and those GitHub repositories will be published via Zenodo. This way, they will be publicly available to anyone for re-use under an open licence. They will be also assigned a Digital Object Identifier (DOI), to make them citable and persistently available.

6. Data management costs

6.1 What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

4TU.ResearchData is able to archive 1TB of data per researcher per year free of charge for all TU Delft researchers. We do not expect to exceed this and therefore there are no additional costs of long term preservation.

The dedicated research software engineer (see the project proposal) will help us to ensure that the published data will be FAIR.