




TECHNISCHE
UNIVERSITÄT
WIEN

Data management plan (DMP)

Retrieval and Analysis of ASCAT and ISMN Data for Soil Moisture Estimation

TU Wien

Version	Effective date	Description of document/changes
1.0	15/04/2023	First version of DMP – created for start of project
1.1	10/05/2023	Second version of DMP – adjusted to fit FAIR principles more accurately
1.2	13/05/2023	Third version of DMP – finalized the document and applied the requirements and repositories

Level of distribution		This DMP is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). DOI: 10.5281/zenodo.7933523
-----------------------	---	--

Project details

Contact person (responsible for data management and DMP)	Aymeric Hollaus, e1508903@student.tuwien.ac.at , ORCID iD: 0009-0009-8783-596X
Contributors	Christian Gäckle, e12010963@student.tuwien.ac.at, ORCID iD: null Artjola Ganellari, e12046001@student.tuwien.ac.at, ORCID iD: null Aymeric Hollaus, e1508903@student.tuwien.ac.at , ORCID iD: 0009-0009-8783-596X, Data Manager
Start date	2023-03-25
End date	2023-05-13

List of acronyms

DMP	data management plan
RDM	research data management
ISMN	international soil moisture network
ASCAT	advanced scatterometer
SM	soil moisture
CDR	climate data record

Content

INTRODUCTION	4
Science Europe practical guide, FAIR data	4
Relevant Policies and Guidelines	4
Publishable Summary	4
1. DATA DESCRIPTION	5
1a Lists of datasets that will be reused or produced	5
1b Data generation and reuse	5
2. DOCUMENTATION AND DATA QUALITY	6
2a Data organisation, metadata and documentation	6
2b Data quality control	6
3. STORAGE AND BACKUP DURING RESEARCH PROCESS	6
3a Storage and backup facilities	6
3b Data security and protection of sensitive data	7
4. LEGAL AND ETHICAL REQUIREMENTS	7
4a Personal data	7
4b Intellectual property rights and ownership	7
4c Ethical issues	7
5. DATA SHARING AND LONG-TERM PRESERVATION	8
5a Data publication and access conditions	8
5b Long-term preservation and deletion of data	8
6. RDM RESPONSIBILITIES AND RESOURCES	9
6a RDM-roles and responsibilities	9
6b Resources	9

Introduction

Science Europe practical guide, FAIR data

A DMP is a structured document that keeps record of what research data is created and what happens to that data during and after a project. It helps with planning the research process and defining responsibilities in a research project involving several researchers or institutions.

For writing this DMP, we followed [the recommendations of Science Europe](#) as they reflect the guidelines agreed upon by the major funders in Europe.

To make our data FAIR, they generally will be treated according to the following criteria:

- We will make our data findable, by uploading it to a data repository that provides a persistent identifier and adding relevant metadata.
- We will make our data accessible by providing open access to data, wherever possible. In cases, where open access is not possible, we will provide meaningful metadata plus contact information for access requests.
- We will make our data interoperable by providing and describing data in a way that is common within our domain by using the same file formats, schemas, and vocabularies. We will provide good documentation for all our datasets.
- We will make our data reusable by adding metadata and comprehensive README files to all published datasets. The descriptions include details on the methodology used, analytical and procedural information. In case of publication, licenses for code and data will always be assigned and clearly marked.

Relevant Policies and Guidelines

- TU Wien Policy for Research Data Management: <https://www.tuwien.at/index.php?eID=dms&s=4&path=Directives%20and%20Regulations%20of%20the%20Rectorate/Policy%20for%20Research%20Data%20Management.pdf>
- TU Wien Code of Conduct – Rules to Ensure Good Scientific Practice: <https://www.tuwien.at/index.php?eID=dms&s=4&path=Directives%20and%20Regulations%20of%20the%20Rectorate/Code%20of%20Conduct%20E2%80%93%20Rules%20to%20Ensure%20Good%20Scientific%20Practice.pdf>
- Directives and Regulations of the TU Wien Rectorate: <https://www.tuwien.at/en/tu-wien/organisation/central-divisions/data-protection-and-document-management/directives-regulations/>
- TU Wien Data Protection: <https://www.tuwien.at/en/tu-wien/organisation/central-divisions/data-protection-and-document-management/data-protection-at-tu-wien>
- European Commission's document on Ethics and Data Protection: https://ec.europa.eu/info/sites/info/files/5_h2020_ethics_and_data_protection.pdf

Publishable Summary

This DMP outlines a project focused on retrieving and analyzing satellite-based ASCAT data and in-situ ISMN data for soil moisture estimation. By leveraging the H-SAF ASCAT SM CDR and ISMN data, the project explores the spatiotemporal dynamics of soil moisture and includes the calculation of Vegetation Optical Depth (VOD). Through the analysis of these datasets, the project aims to enhance our understanding of soil moisture dynamics and their implications for various applications such as agriculture, hydrology, and climate modeling. Following the Science Europe guidance and utilizing the TU Wien DMP Tool layout, the project ensures robust data management practices for reproducibility and facilitates the sharing and reuse of its findings.

1. Data description

1a Lists of datasets that will be reused or produced

Produced datasets

dataset ID	title	type	format	estimated volume	contains sensitive data
P1	experiment_notebook	Software applications	ipynb	1 MB	no
P2	matched_df	Structured text	csv	500 KB	no
P3	matched_insitu_ascat	Image	png	150 KB	no
P4	ascat_data	Dataset	zip	25 GB	no

Reused datasets

dataset ID	title	PID or source	rights (e.g. license)	contains sensitive data
R1	International Soil Moisture Network (ISMN) Data	https://ismn.earth/en/	Creative Commons Attribution 4.0 License	no
R2	H120 - Metop ASCAT SM CDR 12.5 km sampling	https://hsaf.meteoam.it/	EUMETSAT HSAF SAF Data Policy	no

1b Data generation and reuse

Methods and software used for data generation and reuse

To generate the research data, we will be using two main sources: ISMN data and HSAF data. The ISMN data can be accessed by registering at the ISMN website and downloading in situ soil moisture data. This data will serve as the reference data for validation purposes. For the original experiment, a site of interest was selected for our ISMN data, Rietholzbach, ensuring it had temporal coverage within the range of the H-SAF data (2007-2021), which covers at least three years for calibration and validation, and has data available in the topsoil. Once we download the in-situ data, we will upload the .zip file to our GitHub repository and extract said data in the source code (Jupyter notebook). The original HSAF data, on the other hand, can be downloaded from the EUMETSAT HSAF source by registering and accessing their FTP. In this context, we produced a reduced version of this dataset, which is based on the dataset R2, and made it available on Zenodo via the DOI [10.5281/zenodo.7932988](https://doi.org/10.5281/zenodo.7932988). With reduced it is meant that less cell points on the grid are available, so if one would want to look at another location make sure it you have all of the cell points required to do so. Please note that this reduced dataset mainly exists for reproducibility. We also included the necessary files containing information on the grid, soil porosity, terrain, and other auxiliary data in

this reduced dataset. Do note that this extra information, metadata, and ISM data is also included in the GitHub repository, but the Metop ASCAT data will as stated not be included there.

2. Documentation and data quality

2a Data organisation, metadata, and documentation

As of now, the data versioning is limited to the "Last Modified" section of the JupyterLab file explorer. However, we recognize the importance of versioning and plan to implement a more robust system in the future. To ensure that data is structured appropriately, we will create a clear folder structure that reflects the organization of the experiment. We will also use descriptive file names. This will allow for easy identification of specific files and help prevent duplication or confusion. To maintain the integrity of the data, we will establish guidelines for handling and modifying files. All team members will be required to follow these guidelines to ensure consistency and avoid errors or discrepancies. In addition, we will regularly backup the data to prevent loss or corruption. Once a more comprehensive versioning system is in place, we will be able to track changes and maintain a clear record of data versions. This will help us to better understand the evolution of the data and ensure that any changes made are properly documented and recorded.

Explanations of all values and terms used can be found throughout the source code. Due to the source code being a Jupyter notebook, we deemed it to be self-documenting enough to fulfil the requirements by explaining each cell of code in the matching Markdown cells. This will help others to identify, discover and reuse our experiment.

Additionally, we will provide common metadata such as title, description, or data formats when publishing data in open access repositories.

As far as possible, we will use the vocabulary of the geodesy domain and try to make said vocabulary machine actionable.

The data obtained from recognized organizations, such as ISMN and EUMETSAT HSAF, are expected to conform to domain-specific metadata standards and quality control measures established by these organizations. As such, the data from these sources can be considered to adhere to these standards. Detailed information about the data sources, including the metadata, data format, and any preprocessing steps applied by the organizations, will be provided in the README file. This will help others validate the data analysis by understanding the origin, characteristics, and processing steps of the data. For the data created during the research process, detailed documentation and explanation will be provided within the Jupyter cells. The README will include a description of how one can setup and execute the source code, as well as how the data can be accessed since the traditional approach of data access for our kind of data might be tedious for this context. It will also provide an explanation of the values and terms used in the created dataset, ensuring transparency, and facilitating data reuse. By providing comprehensive documentation, we aim to enable others to validate our data analysis and effectively reuse the data for further research or related projects.

2b Data quality control

Peer review of data was and is being done to ensure the data quality and is regarded as our data quality check.

3. Storage and backup during research process

3a Storage and backup facilities

For the duration of the project, storage and backup of data will be ensured by the project manager. The infrastructure of GitHub and Zenodo will be used for this purpose.

P1, P2 and P3 will be stored on GitHub, but will be linked with Zenodo. P4 will be hosted on Zenodo. During the retention period the GitHub repository will be publicly available. P4 will be stored on Zenodo. R1 and R2 have to be accessed and fetched on their dedicated platforms if so desired and or required.

3b Data security and protection of sensitive data

We pay strict attention to compliance with the relevant institutional and national data protection policies listed in the introduction of this document. In this project there will be sensitive data in dataset P1 (Ex1.ipynb). The data used to be non-anonymized (paths, student ID numbers). But during this project it was addressed and will no longer be sensitive data. is taken. Only Project Members will be authorised to access sensitive data.

Access to data during research:

dataset ID	selected project members	all other project members	the public
P1	writing	writing	reading only
P2	writing	writing	reading only
P3	writing	writing	reading only
P4	writing	writing	reading only
R1	writing	writing	reading only
R2	writing	writing	reading only

All incidents will be handled individually by the data manager.

4. Legal and ethical requirements

4a Personal data

At this stage, it is not foreseen to process any personal data in the project. If this changes, advice will be sought from the data protection specialist at TU Wien, and the DMP will be updated.

4b Intellectual property rights and ownership

Legal restrictions on how data is processed and shared are specified in the Terms and Conditions of said organisations. The restrictions relate to datasets R1 (International Soil Moisture Network (ISMN) Data), and R2 (H120 - Metop ASCAT SM CDR 12.5 km sampling). The legal restrictions are based on the following: Usage and sharing of the data is to be for scientific use only, as well as acknowledging and proper citing when required.

Said organisations of datasets have the rights to the produced data and controls access.

4c Ethical issues

No particular ethical issue is foreseen with the data to be used or produced by the project. This section will be updated if issues arise.

5. Data sharing and long-term preservation

5a Data publication and access conditions

As far as possible, obtained datasets will be published in repositories. Details on access conditions, reuse licenses, reasons for restrictions, etc. are collected in the table below.

dataset ID	access conditions	restrictions / embargo reasons	estimated publication date	location for publication (repository)	PID	license
P1	Public	Educational purposes	2023-05-13	GitHub	10.5281/zenodo.7933528	CC BY 4.0
P2	Public	Educational purposes	2023-05-13	GitHub	10.5281/zenodo.7933528	CC BY 4.0
P3	Public	Educational purposes	2023-05-13	GitHub	10.5281/zenodo.7933528	CC BY 4.0
P4	Public	Educational purposes	2023-05-13	Zenodo	10.5281/zenodo.7932988	CC BY 4.0

The repositories used in this project are described in the following paragraph:

GitHub is a version control and collaboration platform for software development. It serves as the central repository for the project's source code, enabling efficient code management, version control, and collaboration among team members.

Zenodo is a research data repository. It is used to store and share research outputs, including datasets and software. In this project, Zenodo is utilized to store the ASCAT data and any extracted data from the Zenodo dataset. It assigns a unique digital object identifier (DOI) to each item, ensuring proper citation and long-term accessibility of the project's data outputs.

Methods or software needed to access and use data: Python, Jupyter Notebook, and access to various libraries. Additionally, a functional email is required to sign up to the various sites for data access!

Description of protocol to access restricted data: The stakeholders will be informed on how they can access the data within the README of the GitHub repository.

5b Long-term preservation and deletion of data

dataset ID	location for long-term storage	minimum retention period	foreseeable research uses and/or users
P1	GitHub	1 year	Members of the scientific community of the TU Wien (including students)
P2	GitHub	1 year	Members of the scientific community of the TU Wien (including students)

dataset ID	location for long-term storage	minimum retention period	foreseeable research uses and/or users
P3	GitHub	1 year	Members of the scientific community of the TU Wien (including students)
P4	Zenodo	1 year	Members of the scientific community of the TU Wien (including students)

6. RDM responsibilities and resources

6a RDM-roles and responsibilities

As this experiment was done with all contributors originally the roles were split between. Due to this being an experiment for the sake of data management, only the contact person and person responsible for data management and this DMP is undertaking the RDM roles and responsibilities.

6b Resources

There are no costs dedicated to data management and ensuring that data will be FAIR.

Additional description:

It is to be noted that this experiment is derived from an assignment done in regard to the Data Retrieval in Earth Observation course at the TU Wien. Therefore the original code and approach stems from the professors and researchers of said course, which are: Vreugdenhil Mariette and Wagner Wolfgang. For the sake of this experiment and assignment the code and approach have been tweaked and adapted, nonetheless this should be kept in mind. The additional members listed above are simply colleagues that helped me when working on the original assignment and are simply here for completions' sake. Lastly some characteristics were tweaked here to fit the assignment specific requirements which aren't FAIR standards from the Science Europe organization, but since it is in this course context it was deemed as okay to be done as such (e.g., 1 year retention period).