TECHNISCHE UNIVERSITÄT WIEN

Data management plan (DMP)

CO2Matters

EOSCSecretariat.eu

Version	Effective date	Description of document/changes
1.0	12/05/2023	First version of DMP

Level of distribution		 DOI: [10.5281/zenodo.7928792]

Project details

Project Coordinator Principal Investigator	Dominik Granig, e11708471@student.tuwien.at.at, ORCID iD: 0009-0009-8898-5361
Contact person (responsible for data management and DMP)	Dominik Granig, e11708471@student.tuwien.at.at, ORCID iD: 0009-0009-8898-5361
Contributors	Dominik Granig, e11708471@student.tuwien.at.at, ORCID iD: 0009-0009-8898-5361, Project Leader
Start date	2023-03-26
End date	2023-05-13

List of acronyms

DMP	data management plan
RDM	research data management
GISTEMP	Global Land-Ocean Temperature Index
GCAG	Global component of Climate at a Glance
DOI	digital identifier of an object

Content

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%20%E2%80%93%20Rules%20to%20Ensure%20Good%20Scientific%20Practice.pd
- 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
- Other (e.g. from a project partner)

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	CO2MattersResults	Plain text, images	CSV, PNG	5 MB	no

Reused datasets

dataset ID	title	source	rights (e.g. license)	contains sensitive data
R1	Annual Global Temperature Time Series	https://datahub.io/core/global-temp	ODC-PDDL-1.0	no
R2	Trends in Atmospheric Carbon Dioxide	https://datahub.io/core/co2-ppm	ODC-PDDL-1.0	no

1b Data generation and reuse

Methods and software used for data generation and reuse

Both the temperature anomalies and CO2 levels are time series. Temperature data is available from 1880 - 2016 in yearly resolution. CO2 level data is available from 1958-03-01 - 2018-09-01 in monthly resolution. To load and transform the data, a python script which utilizes pandas is being used. The main transformation is changing the resolution of the CO2 Levels to also be in yearly format and then trimming both series to start from 1959 so they can be compared against each other.

2. Documentation and data quality

2a Data organisation, metadata and documentation

The output data includes a csv with 4 columns (Year, Temp Anomalys, Average CO2). As the input of the experiment is only historic data, which does not change also the output will stay the same and no version handling is needed.

As there are no domain specific metadata standards applicable, we will provide a README file with an explanation of all values and terms used at project level. This will help others to identify, discover and reuse our data.

Additionally, we will provide common metadata such as title, description or keywords when publishing data in open access repositories. In such a case, we will follow the default template provided by the repository, such as Data Cite Metadata or Dublin Core.

A far as possible, we will use controlled vocabularies for our data to allow inter-disciplinary interoperability and machine-actionability.

Each step of the python script is well documented in the code. The python version was 3.9 and following the standard python guidelines all versions of external packages used are included in a requirements.txt.

2b Data quality control

Both datasets used are already Certified & audited by datahub.io.

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 OneDrive will be used for this purpose.

P1 (temp_anomaly_co2.csv), R1 (Annual Global Temperature Time Series), R2 (Trends in Atmospheric Carbon Dioxide) will be stored in OneDrive cloud as well as a git repository backed up on GitHub.

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

Access to data during research:

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

All incidents will be handled individually by an incident response team that is maintaining the affected service.

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 (Verena Dolovai), 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 PDDL-LICENCE. The restrictions relate to datasets R1 (Annual Global Temperature Time Series), and R2 (Trends in Atmospheric Carbon Dioxide). The legal restrictions are based on the following: ODC-PDDL-1.0 Datahub.io controls the access and distribution of dataset R1 and R2.

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 / DOI	license
P1	Open	none	2023-05-13	TU Wien Research Data	10.70124/4s85p-g0w08	CC BY 4.0

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

TU Wien Research Data is an institutional repository of TU Wien to enable storing, sharing and publishing of digital objects, in particular research data. It facilitates the funders' requirements for open access to research data and the FAIR principles by making research output findable, accessible, interoperable and re-usable. This service is developed by the TU Wien Center for Research Data Management and hosted by TU.it. https://researchdata.tuwien.at/

Data will not be updated after the 13th of March 2023.

5b Long-term preservation and deletion of data

dataset ID	location for long-term storage	minimum retention period (≥ 10 years)	foreseeable research uses and/or users
P1	TU Wien Research Data	10 years	Students and general public

5c Source Code and Data access

The source code is publicly available on GIthub, as indicated in the following table. Running the experiment users need to have python installed and all the required packages in requirements.txt. The resulting data is in CSV format and thus can be freely used for further investigation.

dataset ID	location for long-term storage	minimum retention period (≥ 10 years)	PID / DOI	foreseeable research uses and/or users
C02Matters	Github	1000 years	10.5281/zenodo.7928379	developers and researchers

6. RDM responsibilities and resources

6a RDM-roles and responsibilities

The PI will direct the data management process overall, with the research assistants responsible for ensuring metadata production, day-to-day cross-checks, back-up and other quality control activities are maintained. The lead country researchers will be responsible for routine supervision of the dataset development.

6b Resources

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