How to write a data management plan (DMP)

Version: 1.6

Date: 05 Oktober 2023

TU Wien

Center for Research Data Management

Favoritenstrasse 16, 1040 Vienna

research.data@tuwien.ac.at

Twitter: @RDMTUWien

Content

INTRODUCTION

What is a DMP?

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. You can find more information on DMPs, the TU Wien DMP Tool, and funders' requirements at https://www.tuwien.at/en/research/rti-support/research-data/rdm-infos-tips/dmp.

Why this guide?

Given the fact that many of the addressed researchers/data officers at TU Wien are facing the task of producing a DMP for the first time, the Center for Research Data Management provides TU Wien specific guidance on how to write a DMP. This "how to" is particularly useful if your funder or funding programme (e.g. Horizon) does not provide a mandatory DMP-template. In this guide, we have decided to follow the recommendations of Science Europe because these reflect the guidelines agreed upon by the major funders in Europe.

Who we are

The TU Wien Center for Research Data Management is a central contact point for questions regarding the handling of research data along their life cycle. We are composed of a multidisciplinary team that provides information on suitable infrastructures, services as well as on organisational, legal, and ethical framework conditions. More information at: www.tuwien.at/en/researchdata.

We are developing technical RDM tool and services like the institutional data repository TU Wien Research Data, DBRepo, a repository for databases, and the TU Wien DMP Tool, that helps you to create a DMP for your project. You are welcome to follow our news to learn about new developments.

How to handle this document

Below, you will find chapters on different topics dealt with in a DMP. Each section contains some background information followed by guiding questions (taken from Science Europe's Practical Guide to the International Alignment of Research Data Management, 2018) and a collection of possible answers. The text blocks are either our suggestions or taken from real DMPs, some also from DMPs created by TU Wien researchers. It is important for us to be as close as possible to the requirements of our own institution.

Please note: The given text fragments are **examples only** and your answers may vary a lot depending on the nature and stage of your research project. The text blocks may serve as an **inspiration** for your own text. **Please select only appropriate statements and adapt them to your specific situation.** The collection is neither complete nor does it fit all disciplines or project conditions.

How we want to improve our services

We are constantly working on increasing our collection of good text examples. Therefore, we would appreciate receiving more actual samples from you. If you are willing to share your DMP with your colleagues at TU Wien, please send us your DMP once it has been accepted by the funder. You may as well publish your DMP (e.g. in TU Wien Research Data) and send the link (DOI) to us: research.data@tuwien.ac.at. Please also contact us under this address if you wish to have your DMP draft checked before submitting it to the funder.

THE SECTIONS OF A DMP

0 General information

Providing some general information helps you and others to identify the document and to keep track on the status of the DMP.

Thus, we suggest that you create some kind of cover page with basic information, like title ("Data Management Plan (DMP) for Project XY"), version/date, license and contributors. The license is optional but advisable as it is recommended to register your DMP as a non-restricted, public deliverable that is openly accessible, unless there are reasons to restrict it. In this case, a simple information about the level of distribution like "First version, for internal use only" would be sufficient.

Example for keeping track of DMP versions:

Version	Effective date	Description of document/changes
1.0	XY/XY/XYXY	First version of DMP – e.g. created for start of project, or month X of project as a project deliverable, etc.
1.1	XY/XY/XYXY	Second version of DMP, e.g. prepared for midterm review
1.2	XY/XY/XYXY	

Example for stating the license you assign to your DMP (if applicable):

Level of distribution		This DMP is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). DOI: [xxx]

Example for providing information on project and contributors:

Project Coordinator Principal Investigator	Name, affiliation, e-mail address, IDs (e.g. ORCID iD for person, ROR for institution)
Data officer responsible for data management and DMP	Name, affiliation, e-mail address, IDs (e.g. ORCID iD for person, ROR for institution)
Project title and acronym	
Project abstract	
Start and end date of project	
Funding programme/ grant number	

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 • 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 • Other (e.g. from your project partner)

1 Data description and collection or reuse of existing data

Good to know

Research data comes in all forms and sizes, from different sources and with varying potential for reuse. You can find a definition for research data in the Policy for Research Data Management (RDM) at the TU Wien.

In a DMP you are asked to describe both: existing data you are reusing and data you are generating. You should list the types, formats (often reflected by the filename extension) and estimated volume of your data as well as methodologies or software used for data collection or production. Plus, in case of existing data, the source (data publisher), any constraints on reuse and an explanation of how the data provenance will be documented. If applicable, explain why new data must be collected, rather than reusing existing data.

data types

- text documents (doc, odf, pdf, txt, etc.)
- graphics/images (jpeg, svg, png, gif, tiff, etc.)
- spreadsheets (xls, ods, csv, sas, stata, spss, etc.)
- structured text (html, json, tex, xml, etc.)
- video (mpeg, avi, wmv, mp4, etc.)
- databases (MS Access, MySQL, Oracle, etc.)
- software applications source code (css, JavaScript, Java, etc.)
- audio (mp3, wav, aiff, ogg, etc.)
- configuration data (ini, conf, etc.)
- etc.

sources: generated data

- observations
- laboratory
- field instruments
- experiments
- simulations
- compilations
- sensor data
- observational data
- questionnaires
- etc.

sources: existing data

- GIS data
- Statistik Austria
- data center
- repositories
- etc.

Remember that you will have to refer to the data described here when answering questions in further sections. Therefore, **we suggest that you simply collect the information on your data in a table** and add an ID for references within the DMP (see example below). This will give you and others a clear overview and help you later to distinguish between different datasets.

Formats

Open and standard formats are preferred as they facilitate sharing and long-term reuse of data. If you are planning to convert your data into open or standard formats, you should describe your strategy.

Nevertheless, there may be reasons for the use of other formats which should be named in the DMP. Motivation could be:

- widespread usage within the research community
- generated by the software or equipment that will be used
- staff expertise within the host organisation
- standards accepted by data repositories.

You should also consider machine readability of the data (and documentation). For example: a table with numbers in PDF is not considered machine-readable, the same data in CSV, JSON or even XLS is.

Volumes

Details on the (expected) volumes can be expressed in storage space required (bytes), or in numbers of objects or files.

Guiding questions and inspirations for your answers

1a	What data (for example the kind, formats, and volumes), will be collected or produced?

ID	name	type	format	volume	reused/ produced	comment
1						
2						
3						
4						
5						

We will **produce** the following data:

- Excel files containing laboratory measurements in XLS format. Their size will not exceed 100 MB.

We will **reuse** the following data:

- Patient interviews published by XYZ that are publicly available at https://doi.org/...

  They are available as PDF documents and their size is about 50 MB.

This project produces aggregated datasets in CSV format (10 x ~10 MB) that contain data points that combine alcohol consumption data with UFO sighting data and a correlation plot of these in PNG format (<10 MB).

The project reuses two external CSV datasets with statistical data:

1. Dataset: Alcohol Consumption: OECD (2018), Alcohol consumption (indicator). DOI: 10.1787/e6895909-en, 112 KB
2. Datenset: Ufo Sightings, Sigmond Axel. (2014). Ufo-reports. https://github.com/planetsig/ufo-reports, 13 MB (accessed on 22 March 2018)

Research outputs:

- merged and pre-processed data generated from the initial collected data
- Jupyter notebook including the CRISP-DM workflow and the final results of prediction and analysis

---

Files are stored in open, standardized formats as far as possible. The formats PDFA, CSV, MPEG-4 (audio track WAVE) and if necessary, TIFF are used for this purpose. Where conversion to an open format is not possible, original formats are saved.

The intermediate data can be hundreds of GB, which we will not store. We envision the total volume of all the final result data, input data sets, source codes, etc. to be below 50GB, which will be stored and backed up as per the procedure, explained later in this document. Please note that, depending on the final developed framework at the end of the project and the involved libraries, their corresponding licensing and versions of the operating system used, we may also provide a virtual machine with the necessary input data sets, etc. to facilitate reproducible results and research. In that case, some additional volume of 10-20GB may also be required, which is easily manageable considering the available resources at our institute and TU Wien's central IT infrastructure.

As the data is very domain-specific and experiment-specific, interoperability is low. However, standard file formats for which many parsers exist (JSON, CSV) are used, making it possible to import data in many languages, etc. The tools are also cross-platform and open source and will therefore not only work on Linux and my development computers but everywhere if needed.

Source code of software tools (in a high-level language like C/C++, Python, and MATLAB) as well as the hardware designs (in a hardware description language like VHDL or Verilog) for different analysis, modelling, and optimization techniques that will be researched in the scope of the proposal for reliability analysis, modelling and mitigation including both soft errors and aging-related reliability issues.

No new data will be produced or generated by the project.

1b	**How will new data be collected or produced and/or how will existing data be reused?**

Datasets #1 to #5 will be produced in a laboratory by performing XYZ.

Datasets #1, #2 will be collected in the lab, #3, #4 are results of simulations done using VSC.

The research data will be generated through extensive simulations. One key method to analyze the reliability of a program and architecture is fault-injection experiments, for which we will perform Monte-Carlo simulations. For the program reliability analysis, we will leverage a (functional/cycle-accurate) instruction set simulator like Gem5 and Sniper. Besides using machine learning applications like deep neural networks developed using the open-source Pytorch framework, we will rely on open benchmarks for evaluation like using PARSEC and MediaBench or the recent MLPerf benchmarks. For the hardware-level reliability evaluation, we also plan to perform RTL- or Gate-level reliability analysis (e.g., using low-level fault injection experiments) using the Synopsys and Cadence tools.

Generated data originate from analytical equipment that will be used during the duration of the project, and of different microscopes (for the images). Specific software will be used to process the raw data, to conduct statistics, to create graphical representations, or to make drawings.

Focus groups will involve two researchers and be conducted in the vernacular. Whether recorded or not, the event will be transcribed or documented using agreed formats and standards for handling the issue of multiple voices, interruptions, labelling of participatory and visual activities, and so on. All transcripts will be in Microsoft Word. Focus group and interview transcripts will be coded in NVivo or a qualitative software suited to the different languages; the most appropriate software for a comparative multi-language study has not yet been identified.

Focus groups are organised and surveyed throughout Austria. The answers are stored as video recordings and subsequently transcribed. The evaluation of the answers is done by MAXQDA. Excerpts from the videos will also be used for teaching and further education. Existing data will also be used: A secondary analysis of the statistics on … of the Statistik Austria will be carried out. The data are evaluated with the help of the statistics program R.

We will develop models based on literature review and prototype development.

2 Documentation and data quality

Good to know

In a DMP, you are asked to provide information on how data will be organised during the project, which data quality control measures are used, what documentation will be prepared and made available to enable reuse of data, which metadata will be provided, and which metadata standards are used.

FAIR principles

Some of the questions in this section focus on the FAIR principles. Following these principles will help you to make your data findable, accessible, interoperable, and re-usable (which is, by the way, a requirement from TU Wiens RDM policy). Briefly explained:

- datasets are findable if they are published in a suitable data repository and have a permanent and unique identifier (e.g. DOI)
- data is accessible if it is clear how to get access to it, even when it is closed
- interoperable means that other researchers know how to interpret the values (clear metadata) and that the data is published according to standards used within the given community, e.g. interviews are available as MP3 files and not vendor specific files that require specific hardware/software for playback

- to be reusable, data needs to be companied be a detailed description and a license for clear information on reuse rights and conditions.

You can find more information on the FAIR principles on our website.

Data organisation

Data organisation refers to the arrangement of data for retrieval. By improving data organisation, you improve the findability of your data, for yourself and for other data users. Additionally, clear structures and clear folder systems make it easier for you to manage your data, which plays an important role with regard to simple and reliable access control for sensitive data, for example.

In the DMP, you are asked to describe how data will be organised during the project.

data organisation methods

- naming conventions
- clear storage/folder structures
- version control
- etc.

Documentation

The documentation outlines the research process and ensures integrity, understandability and transparency of the data collection process and facilitates correct interpretation. This may include information on the methodology used to collect the data, analytical and procedural information, definitions of variables, units of measurement, and so on. Check: Has the software needed to access the data been sufficiently documented? Is it possible to include the relevant software, e.g. in open source code?

In the DMP you must clearly outline how this information will be captured and archived with the data.

documentation

- README text file
- lab journal, lab notebook
- methodology section of papers

- data collection context: structure and organisation of data fields, variable names and description
- methodology of data collection: definition of codes for access to data and classification schemes
- algorithms for transformation and processing of data
- method reports, instruments of data collection, codebook, program code for data processing, data code for analysis
- theoretical description of the measurement data (e.g. with Assyst, LabVIEW)
- informed consents
- etc.

Metadata

Metadata are data about data. They are needed to easily identify and discover data. Metadata standards are useful to structure metadata in a commonly agreed way. You can find more information on metadata on our website.

The easiest way to provide clear metadata is to use discipline specific metadata standards. If you are not aware of any metadata standards used in your domain, you should check the following websites for help:

- FAIRsharing
- re3data (filter: metadata standards)
- Digital Curation Centre (DCC)

If there is no suitable metadata standard for your data, you should indicate this in the DMP and use some kind of documentation (e.g. README file, see above) to provide the relevant metadata. Make sure that data, concepts, columns, labels, etc. can be understood by people outside your research group - and outside your research domain as well.

Repositories provide common metadata templates that are domain independent. You should adhere to the repository metadata schema and provide as much information as possible when loading up data (title, description, keywords, type of publication, etc.).

If there is a metadata standard for your domain

- use it and declare it,
- find a repository that supports this standard, e.g. by using re3data
- if there is no repository, use a general-purpose repository. In such a case you will have to provide common metadata such as title, description, etc. plus a README file as mentioned above.

metadata standards

- Dublin Core for semiotic entities
- Data Cite (data information mandatory for getting a DOI)
- SKOS to describe knowledge organisation
- EXIF for camera settings (photos)
- EML for the ecological sciences
- etc.

Data quality

In your DMP, you should clearly describe how the consistency and quality of your data collection will be controlled and documented.

data quality control processes

- repeat samples or measurements
- calibration
- standardised data capture
- data entry validation
- peer review of data
- representation with controlled vocabularies
- etc.

Guiding questions and inspirations for your answers

2a	**What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?**

FAIR principles

The data 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 (see section xyz).
- 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 data sets. 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.

The required data to fulfil the FAIR criteria ranges from general information like author, title, type of data, etc. to very specific information such as configuration settings, software versions, build details, e.g., for applications, simulators, and other software implementations. We will make all the research results along with the corresponding configurations open for reproducible research, while taking care of the 3rd-party license constraints from XYZ and other tool and library providers. Since we will rely on open-source tools as much as possible and use open benchmarks, we do not expect any issues with reusing the data. Towards the end of the project, we plan to prepare a virtual machine package with all the open-source simulation environments, executing applications and work package developments.

Data organisation

Filenames will follow the projects naming convention as defined in document ABC.

The filenames will include a timestamp of creation.

The Embedded Computing Systems Group has an in-house GitLab platform to handle proper planning, management, versioning, monitoring of and collaboration on different R&D efforts. We will use this platform, where for each work package, we will create separate repositories, and versioning will be done at appropriate milestones, e.g., once a technique is mature and tested to give benefits over state-of-the-art, or once a developed technique has passed the validation tests.

The respective work package leader will handle the structure and versioning of the research data. For the whole project a standardized folder structure and naming conventions will be in place.

A version control is automated. File naming is done according to the following standard:
[focus group]_[location]_[YYYYMMDD].mp4

Metadata and documentation

There are no domain specific metadata standards applicable. We will provide a README file with an explanation of all values and terms used next to a file with 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.

When taking pictures, we will record the camera settings by using the metadata standard EXIF.

We will include README-files at file-level, dataset-level, and project-level. The files will contain data inventories and important contextual information such as the software used to collect/process the data and any assumptions made during analysis.

The methods report will outline the data collection and processing activities. The code that was used to gather information online will be supplied. A codebook will provide an overview of the variables.

Each study-folder will contain a short word document explaining the project title, date, people involved, the main aim, methodology, main results, and a short conclusion. This word document will provide the page number of the Lab Book, where the methodology and any other important information were noted by hand by the student. All this documentation will be written in English.

All the source codes will be properly documented in the design and source files. The documentation of the code will be extracted via tools like doxygen and can be stored in html or pdf formats. Moreover, there will be a quick How-To guide for regular users provided as text or html files. All the constraints files will also be provided in their relevant format like text or xml/json.We will make source code findable, by uploading it to an open data repository that provides a persistent identifier. We will make it accessible by providing open access, wherever possible. We will make it interoperable by following current coding conventions and technologies. Our license will allow access and reuse of data. For closed data, we will provide descriptive metadata.

Metadata are created by xyz according to the DDI standard. Keywords will be assigned according to the subject-specific thesaurus xyz. Additional documentation of the research data is also planned. The following documents will be created:

- transcription manuals
- focus group guidelines
- QDA files
- R-syntax
- declarations of consent
- anonymisation measures

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

2b	What data quality control measures will be used?

Data quality checks will be done, e.g. checks of consistency of labels, logical errors in the data, data curation, and version control.

Source code review and refactoring will be done.

A data quality (DA) dimension will be addressed which consists of accuracy, completeness and consistency or measures/indicators. Some of the data quality issues can be “missing data”, “incorrect data”, “irrelevant data” etc.

Surveys: Focus groups and interviews will always involve two researchers. Quality control for the qualitative data collection will be assured through refresher focus group discussion training during research design workshops and to junior researchers, where appropriate. Where translations are undertaken, quality will be assured by one other researcher fluent in that language checking against the original recording or notes.

3 Storage and backup during the research process

Good to know

During the project, you need to make sure that your data is stored safely and backed up as required. You can find an overview of TU.it services for storing and sharing data during the research process on our website.

In the DMP, you are asked to clearly describe where your data will be stored and backed up (at least two separate locations) and how often backups will be performed. It is recommended to use TU.it storage services or robust, well managed storage with automatic backup of your partner institution. If institutional storage cannot be used for (parts of) your data, explain why and describe where and how data will be stored and backed up instead. Storing data on laptops, stand-alone hard drives, USB sticks etc. should not be an option.

In section 3b of your DMP you should mention the TU Wien data protection policy and describe the measures planned in your project to ensure data security and protection of sensitive data during the research process. Potential topics are:

- data recovery in the event of an incident
- access management and access control (particularly in collaborative partnerships); who will have access during the research process?
- protection of sensitive data: main risks during storage and transmission and how they will be managed (additional security measurements)

data storage during the project

- TUproCloud
- TUfiles
- institute’s server with automated backup
- etc.

data security

- physical security
- network security
- security of computer systems and files
- etc.

sensitive research data can involve

- trade secrets
- national security information
- 'Special categories of personal data': When people are involved, personal data can reveal:
  - racial or ethnic origin
  - political opinions, religious or philosophical beliefs
  - genetic, biometric or health data
  - sex life or sexual orientation
  - trade union membership
- etc.

Guiding questions and inspirations for your answers

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

During the project, the data is stored on TUfiles, a central and redundant network drive with daily backups and regular snapshots provided by TU.it.

Storage and backup will be ensured for the duration of the project by the project manager in cooperation with the responsible representative of TU.it. The infrastructure of TU Wien will be used for this purpose. The research data will be shared with partners on the project's TUproCloud, a sync&share service provided by TU.it. Only authorized staff members and project partners have access.

During survey data collection, data is stored with the online survey tools LimeSurvey in Germany, according to a GDPR compliant data processing contract between the TU Wien and LimeSurvey GmbH concerning personal data. Data Protection at TU Wien provides for a template [*Auftragsverarbeitervertrag*].

3b	**How will data security and protection of sensitive data be taken care of during the research?**

Project members will not store sensitive data on computers in the lab or external hard drives connected to those computers. They will not carry data with them (e.g. on laptops, USB sticks, or other external media). All data centers where project data is stored have sufficient certifications. All project web services are addressed via secure http (https://...). Project members have been instructed about both generic and specific risks to the project.

Voice recordings are made. Although any connections to the speakers are deleted, they are still sensitive data. For analysing the transcripts will be safely stored on the TUproCloud, where only people who are directly involved with the analysis of the samples have access. The original voice file will be deleted after the transcription.

The project produces no sensitive data.

4 Legal and ethical requirements, codes of conduct

Good to know

Legal aspects

You can find information on the important legal aspects of research data management as well as links to relevant policies, codes of conduct and contact persons for legal advice at TU Wien on our website. Data ownership, personal data and granting of licenses for reuse of your data are the main topics in this section.

In the DMP, you need to show that requirements linked to the collection/use and lawful processing of personal data (incl. anonymisation, access management, data sharing issues etc.) have been considered. Give a clear statement if you are not dealing with personal data in your project.

Other legal issues that need to be dealt with in the DMP are intellectual property rights/ownership and rights to control access (differentiate by subsets if necessary), especially in multi-partner projects. Clearly explain how these matters are explained in the consortium agreement.

Please note: Your answers to the legal issues will be very different depending on your project and your requirements. Copy/paste of ready-made sentences could involve risks. Please be aware that at TU Wien, you can have the service for legal review:

- The Research and Transfer Support Office supports researchers and research groups at the TU Wien in various research and exploitation activities, e.g. patent and license management and the design and negotiation of R&D contracts
- If you have any questions regarding data protection or the processing of personal data within research data management, please refer to the website about data protection at TU Wien.

potential legal issues

- data ownership, IPR
- personal data (GDPR)
- restriction for publication
- embargo periods
- etc.

Ethical aspects

In the DMP, you should consider whether ethical issues can affect how data are stored and transferred, who can access them, and how long they are kept and demonstrate awareness of these aspects and respective planning. It is a good idea to point to relevant national and international codes of conducts and institutional ethical guidelines, and mention if ethical review (for example by a research ethics committee) is required for your data collection.

There may be ethical reasons for fully or partially restricting access to research data. Such reasons include, for example, avoiding risk of harm to people (individuals, small groups, minorities), the environment, or even society. Please note the broad notion of harm which includes besides physical harm, psychological, financial, reputational, emotional harm. If applicable, describe in your DMP, which measures are in place to prevent or mitigate the data-related harm to participants.

Make sure you address all research ethics questions (participation of humans, use of human samples, 3rd country participation in your research, potential dual use/misuse) related to your work in the proposal as well. If you need help identifying possible ethical questions in your research or formulating statements on research ethics in your DMP, please contact TU Wien's Service Unit of Responsible Research Practices.

ethical issues can for example surface in connection with

- experiments and studies that involve people (surveys, prototype tests, interviews, focus groups, etc.)
- research involving human samples such as cells or tissues
- environmentally relevant research that can pose a risk to people themselves and the environment
- exchange of knowledge and technology or research data collection with third countries
- security-related research, where there is potential for misuse of research results
- materials, knowledge and technology with dual use potential
- etc.

We recommend the following guidelines and publications for further information:

- Guidance document: Informed Consent: Good practice recommendations, TU Wien, December 2021
- Ethics and data protection, Directorate-General for Research and Innovation (European Commission), July 2021
  → contains among others explanations of the concepts pseudonymization and anonymization
- The European Code of Conduct for Research Integrity, ALLEA, revised edition 2023, ISBN 978-3-9823562-3-5
- Identifying serious and complex ethics issues in EU-funded research, Directorate-General for Research and Innovation (European Commission), July 2021

Guiding questions and inspirations for your answers

4a	**If personal data are processed, how will compliance with legislation on personal data and on security be ensured?**
Important to know: If personal data is involved in your project, you might need to conclude a "data processing contract" [*Auftragsverarbeitungsvertrag*] to be GDPR [*DSGVO*] compliant. Data Protection at TU Wien provides for a template.	

Personal data will be stored an processed in accordance with the TU Wien data protection policy:… [explain how]

Compliance with applicable legislation will be ensured by providing data protection information, use of TOMs (Technical and Organisational Measures), informed consent forms, encryption/anonymisation of personal data and managed access to the data in consultation with the responsible data protection officer at TU Wien.

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

As the data sets are from the industrial and health sectors the collected data as such will generally not be published (only associated metadata). However, partial releases may occur of anonymized and aggregated data in the course of the project and in coordination with the industry partners and the Research Ethics Coordinator at TU Wien.

Before publication, the data will be pseudonymized to protect respondents' privacy according to the GDPR and Austrian national law.

Only project staff who sign a non-disclosure agreement will be granted access to the raw data. Before publication, the data will be anonymized to protect respondents' privacy according to the GDPR and XY national law.

Commitments to ensure confidentiality between the project partners will be maintained by ensuring datasets XY are not shared; datasets XY are anonymised and details that can be used to identify participants are removed.

Data collection is based on an information sheet/consent form for respondents including all relevant information on data processing. Only respondents who agreed to the informed consent are allowed to proceed with the survey. Also, the identity of participants to the survey is protected and no negative effects of the participation in the survey is foreseen. Only project staff who sign a non-disclosure agreement will be granted access to the raw data.

(Surveys) A letter explaining the purpose, approach, and dissemination strategy (including plans to share data) of the research, and an accompanying consent form (including to share data) will be prepared and translated into the relevant languages.

4b	**How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?**
Please note: Data ownership is a question to be answered individually for each project. There are various possible options: TU Wien, (an)other project partner(s), all project partners together, the funding agency. If in doubt please contact our legal experts at Research and Transfer Support Office, vertragsservice@tuwien.ac.at.	

Ownership

TU Wien will be the owner of the data generated and have the rights to control access. Further details will be covered in the consortium agreement.

Project partner(s) xxx will be the owner(s) of the data generated and have the rights to control access. Further details will be covered in the consortium agreement.

All project partners will jointly own the data generated and have the rights to control access. Further details will be covered in the consortium agreement.

The funding agency will be the owner of the data generated and have the rights to control access. Further details will be covered in the consortium agreement.

Online and archival sources will be cited and clearly acknowledged in the database and research outputs. Permission will be sought from secondary sources to share the findings of the research on public websites.

Licenses

The project partners will publish developed tools, models, output and input data in a user-friendly condensed format, for peer review and non-commercial use. As a default license for foreground IP generated in the project, we plan to use Apache 2 License for the software code and Creative Commons CC-By for the datasets.

Within our research group, we have agreed to the following IPR and licensing conditions:

- Data owner: [name]
- Licenses: license CC-By 4.0 for datasets and license XY for software
- Restriction: for datasets AB and CD until [date]
- Embargos: for dataset XY until [date]

The code will be licensed under the MIT license.

The digital research data obtained will be published Open Access under a Creative Commons CC-BY license, provided that there are no data protection concerns. Further data will be made available with restrictive access. Repository xyz is used for making the data available.

All documentation of the project, including the publications and reports, will be available under the Creative Commons Attribution CC-BY license (https://creativecommons.org/licenses/by/4.0/). The software generated as a part of this project will be made openly available under the MIT license (https://opensource.org/licenses/MIT), with the following license text:

- Copyright <YEAR> <COPYRIGHT HOLDER>
- Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the software.

4c What ethical issues and codes of conduct are there, and how will they be taken into account?

Ethical issues in the project have been identified and discussed with the Research Ethics Coordinator at TU Wien (https://www.tuwien.at/en/research/rti-support/responsible-research-practices). They relate to… and will be resolved by…

Relevant ethical guidelines in this project are…

No particular ethical issue is foreseen with the data to be used or produced by the project.

5 Data sharing and long-term preservation

Good to know

Data sharing occurs during the project (sharing among project partners) and after the project (publishing or sharing on request) and is often closely related to the topics of data storage (see section 3) and long-term preservation. You can find more information on preserving and publishing data on our website.

This section of the DMP focusses on depositing data into repositories - for sustainable data management beyond the lifetime of the project and to make data discoverable and reusable to others. You are asked to show that you will use a trustworthy data repository, indexed in a catalogue, and give information on how long data will be retained. According to the RDM Policy of the TU Wien, data have to be stored for at least ten years but funders' may ask for an even longer period of time.

It is recommended to demonstrate in the DMP that the policies and procedures of the chosen repository (including metadata standards and costs) have been checked. Trustworthy, long-term repositories provide persistent identifiers (e.g. DOI) so that data can be reliably and efficiently located and referred to.

If available, you should use a domain specific repository for your data. Check the following platforms to find out, if there is anything suitable for your area of research:

- re3data.org
- fairsharing.org

Otherwise, we recommend to use TU Wien Research Data, TU Wien's institutional research data repository.

repositories

- domain specific repositories
- TU Wien Research Data: https://researchdata.tuwien.at/
- TU Wien GitLab: https://gitlab.tuwien.ac.at/users/sign_in
- Zenodo: https://zenodo.org/
- etc.

Long-term data preservation usually plays an important role at the end of the project but must be prepared in time - strictly speaking: before the project starts. Generally, it does not make sense to preserve all collected and produced data in all versions. In case of reused data, it is often sufficient to quote the source.

Some datasets must be retained for contractual, legal, or regulatory purposes. On the other hand, there might be data, that need to be destroyed for legal or ethical reasons. This has to be considered and documented carefully. Therefore, it is recommended to define a selection process and describe the data and its associated documentation that will be preserved long-term within the DMP.

You are also asked to explain the foreseeable research uses (and/ or users) for the data and give an estimation in your DMP, when your data will be made available to others. This could be straight away, at the time of an article publication, or postponed (e.g. to protect intellectual property or seek patents). If exclusive or restricted use of your data is claimed, clearly explain why and for how long and confirm that you will make associated metadata and contact information available to others. If it is necessary to restrict data access to certain communities, indicate how and why and explain who will be able to use the data. For potential publication restriction based on ethical reasons see chapter 4.

Briefly indicate whether specific software or tools are needed to access, interpret and reuse the data and consider the sustainability of required software. More details on this topic will be compiled in the documentation. If it is not possible to make specific tool and software available to others, clearly explain why.

Guiding questions and inspirations for your answers

5a (1) How and when will data be shared?

Research data will be made available for reuse in a suitable (and if available certified) repository. Specific details on the repository, persistent identifiers, metadata standards, and licenses are not yet available. The DMP will be updated with that information in a revised version.

Data xyz will not be made available for reuse due to its size. It is unfeasible to deposit the data in a repository in its entirety. Metadata will be made available in TU Wien's research data repository together with contact data for more information.

Tools and software will be made available… [explain how, when and where].

During the project, we will be using TUproCloud to exchange data between partners.

Data underpinning research papers will be made available in the TU Wien data repository TU Wien Research Data at the time of the article's publication.

According to the data processing contract [German: Auftragsverarbeitervertrag] between TU Wien and XYZ, the following personal data will be made available in the TU Wien Research Data repository at the end of the embargo period [month, year].

According to the cooperation contract, exclusive use of the research data will be claimed until one year after the end of the project to protect intellectual property / seek patents. Nevertheless, data underpinning research papers will be made available to other researchers at the time of the article's publication.

Whenever concrete validated results are obtained, the corresponding data (input, output, configurations, simulator state, etc.) will be stored. Moreover, all the regular data will be kept on the group's file server, which is regularly backed up.

5a (2) **Are there possible restrictions to data sharing or embargo reasons?**

At the moment, we are not aware of any restrictions on the reuse of data. If issues arise during the project, we will discuss them with our legal/ethical advisors and update the information in the DMP.

We do not plan to impose any restrictions on the reuse of data generated and published in the course of this project. In cases where pre-existing licensing restrictions prevent full publications, we will provide at least one publicly available example.

The reuse of dataset #3 is restricted due to the intellectual property rights of industry and health sector partners.

5b (1) How will data for preservation be selected?

All data required to reproduce published research results will be archived in long-term storage solutions and made publicly available as far as previously existing license restrictions permit. Data that is outdated or temporary in nature will be deleted at the end of the project.

Research data on which a publication is based, but also other relevant milestone files of the project are stored safely for at least ten years in TU Wien Research Data, TU Wien's institutional data repository. Data for which there is no legal archiving basis will be deleted shortly before the end of the project. The data protection officer of TU Wien is involved in this process regarding personal data. The expected total size of the remaining data is about 100 GB.

According to the Policy for Research Data Management (RDM) at the TU Wien, datasets #5, #6 and #10 will be stored for minimum 10 years.

Target audience: researchers, PhD students, master & bachelor thesis students, potentially also the researchers at industrial and academic research organizations working in the area of dependable computing.

Target audience: researchers in environmental engineering and society as a whole.

5b (2) Where will data be preserved long-term (for example a data repository or archive)?

Data will be stored for at least ten years in accordance with the RDM Policy of the TU Wien.

Tape storage as a long-term backup, as well as Git repositories will be alive.

Data recovery is ensured since all work is done on TUgitLab, TU Wien's institutional GitLab instance.

For long-term archiving of and access to the datasets, the Zenodo repository will be used. For source code, GitHub will be used, where also Zenodo can be used for archiving and creating a persistent identifier. For making the envisioned publications (e.g., journal and conference publications, scientific reports, master/doctoral theses) available, reposiTUm will be used (pre-print, post-print, published version depending on the respective restrictions, if

applicable), which is the institutional repository for scholarly publications managed by TU Wien.

For publication of data, we will be using the following repositories:

- Zenodo to publish project outputs such as reports and data
- the domain specific repository XYZ for X
- we will make GitHub repositories public as well for making code accessible.

Persistent identifier will be assigned by the repositories. With these sharing and storing practices in place, we also support the FAIR principles.

The Jupyter notebook will be published in GitHub together with the documentation. A DOI will be assigned (using Zenodo and GitHub integration). Thus, we ensure the Jupyter notebook is preserved for the long-term by CERN. This will require no additional cost at our side.

5c	**What methods or software tools are needed to access and use the data?**

The cell libraries and other 3rd party data/tools which are covered under a closed license cannot be released. However, we will provide all the technical details, and if other researchers can obtain the same set of libraries and tools, they should be able to reproduce our results. Please note that we will follow the widely-used tools from Synopsys, Cadence and Xilinx, and therefore do not envision any restrictions towards reproducible research in case other research groups have access to the similar tools and libraries.

5d	**How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?**

Persistent identifier will be assigned by the repositories.

The Jupyter notebook will be published in GitHub together with the documentation. A DOI will be assigned (using Zenodo and GitHub integration). Thus, we ensure the Jupyter notebook is preserved for the long-term by CERN.

6 Data management responsibilities and resources

Good to know

It is worth thinking about responsibilities and resources at an early stage of the project or even before the project starts: to calculate additional costs and time requirements in and to make sure that roles and RDM-tasks are assigned in good time and in a transparent and clear way.

Roles and responsibilities

Where possible, you should name responsible individuals when outlining the activities, roles and responsibilities in the DMP. Data management is particularly important in collaborative projects and requires good communication and coordination of responsibilities among partners.

data management activities

- data capture
- metadata production
- data quality control
- storage and backup
- data preservation/deletion
- data sharing/publication
- writing, reviewing and updating the DMP
- day-to-day implementation of the DMP
- coordination of DMP responsibilities across partners
- ensuring that data are FAIR
- etc.

Costs

You are asked to provide a realistic estimation of necessary resources for data management and a clear statement on how these costs will be covered. Please note that some funders/funding programs accept data management costs in grant proposals. You can find further information on RDM-costs on our website.

data management costs

- personnel costs, e.g. hiring a person dedicated to data management in the project
- staff time, e.g. for the preparation of research data for preservation and publication including the relevant documentation
- acquisition of existing data, e.g. from data archives with underlying payment terms
- hardware and infrastructure, e.g. for cloud computing and storage, High Performance Computing, network (in case of special applications); repository charges
- software, e.g. Electronic Lab Notebooks, project management software, licenses (if not offered by TU Wien)
- programming, e.g. for APIs regarding data deposit, visualisation purposes
- legal advice, e.g. for copyright questions, special agreements, consent forms
- etc.

Guiding questions and inspirations for your answers

6a	**Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?**

The [PI] [data manager XY] [Partner XY] will direct the data management process overall, with the research assistant 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.

The [PI] [data officer] is responsible for the secure storage and preservation of the generated digital research data together with the institute's IT representative.

6b	**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)?**

Necessary time resources to prepare the data for sharing/preservation (data curation) have been costed in (see grant application).

Xyz FTE are provided in the project for the appropriate preparation of the research data for preservation and publication. Preserving and publishing the data in TU Wien's data repository TU Wien Research Data is free of charge.

Additional resources will be needed to prepare data for deposit and to cover charges from data repositories. According to the cooperation contract between TUW and XYZ, such costs will be covered by …. [explain how].

No additional resources are needed for data management and storage.

The storage costs of approximately xyz € will be covered by the group's annual budget.

ANNEX

Core Requirements for Data Management Plans

[Source: Science Europe (2018), Practical Guide to the International Alignment of Research Data Management, p 9-10.]

CORE REQUIREMENTS
FOR DATA MANAGEMENT PLANS

When developing solid data management plans, researchers are required to deal with the following topics and answer the following questions:

1. Data description and collection or re-use of existing data

a. How will new data be collected or produced and/or how will existing data be re-used?

b. What data (for example the kinds, formats, and volumes) will be collected or produced?

---

2. Documentation and data quality

a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany data?

b. What data quality control measures will be used?

---

3. Storage and backup during the research process

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

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

---

4. Legal and ethical requirements, codes of conduct

a. If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?

b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

c. How will possible ethical issues be taken into account, and codes of conduct followed?

---

5. Data sharing and long-term preservation

a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

b. How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)?

c. What methods or software tools will be needed to access and use the data?

d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

6. Data management responsibilities and resources

a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

b. 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)?