TAU DMP Guidelines 2020

Motivation

How to write a DMP? - READ THIS FIRST!

- Read all the questions first!
- Use DMP to complement your research plan - avoid overlapping with the research plan!
- **Research plan** describes the scientific, data analysis and methodological processing of data.
- **Data management plan** describes technical and administrative management of data.
- To avoid redundancy, refer to your research plan in your DMP and vice versa.
- Use DMP as a risk evaluation document - show that you can recognize, anticipate and handle the risks related to your data management workflow.
- Reflect DMP on your own research project - Do not copy/paste examples from somewhere else.
- Write only sentences you yourself understand.
- Answer the questions where applicable - If a certain question is not applicable in your case, justify why not.
- Answer at least the main categories - Each sub-question does not need to be answered separately.
- Include background information such as the name of the applicant and the project, project number, funding programme, version of DMP.
- Show your data management and version control skills, for example, consider the file naming of the DMP file.
- Follow the organisation or funder requirements.

Why should you manage your research data and write a data management plan (DMP)?

- It is good research practice!
- You will reduce the risk of losing your data.
- You will be able to anticipate complex ownership and user rights issues in advance.
- It helps you support open access to create productive future collaborations.
- You will meet funder requirements.
- It helps you save time and money.
- Your DMP reflects your managerial skills as a project leader.

In the DMP context, data is understood as a broad term. Data covers all information and material your research results are based on. You can concentrate on the data that is your responsibility.

Your DMP should describe how you will manage data **during the whole research life cycle**. The DMP is a living document that should be updated as the research project progresses.

Your research data management practices should aim to produce reusable data that follows the **FAIR** principles, that is, your data will be Findable, Accessible, Interoperable and Re-usable.

Good luck with your DMP!
1. General description of data

1.1 What kinds of data is your research based on? What data will be collected, produced or reused? What file formats will the data be in? Also, give a rough estimate of the size of the data produced or collected. (Data Description)

Briefly describe here:
- what data will you use or produce in the project?
- in which file formats will the data be in?
- how much data will you have (approximately)
- if you will use or develop special software or code to analyse your data.
- if you use personal or sensitive data (see the guidance after “Tips for best practices”)

Your answer to this question forms a general structure for the rest of the plan. Categorise your data in such a way that you can refer to it later in the plan. For example,

A. Data collected or produced by you or your research group
B. Data collected by other researchers
C. Data from other sources such as registers, statistics, measuring stations etc.
D. Other materials needed to use and understand the data, such as codes, softwares, lab notebooks etc.

The categorisation follows the license policy of your data sets. For example, briefly describe according to which license you are entitled to (re)use the data.

In DMP, describe the required disk space - not how many informants were participating the project. A rough estimation of the size of the data is sufficient - e.g. less than 100 Gb, approx. 1 Tb, or several petabytes.

Tips for best practices

- When categorizing your data, use bullet points for a concise way of presenting
  - data types
  - The file formats (for example, .csv, .txt, .docx, .xlsx, .tiff) used during the research project may differ from those used in archiving the data. List both. The file format is a primary factor in the accessibility and reusability of your data in the future.
  - Favour software and formats based on open standards to enable data reuse, interoperability and sharing.
  - the size of the data sets
  - the software used (especially if the software is coded in your project)
  - other relevant information related to your data sets.

- AVOID OVERLAPPING WITH THE RESEARCH PLAN! Data analysis and methodological issues related to data and materials should be described in your research plan.
Guidance for sensitive and personal data

Specify all dataset types that contain personal, sensitive or confidential data. Identifying the sensitive components of research data is particularly important, as the planning of data management focuses on the identification and management of related risks. If you work with personal data, specify the party serving as the controller.

Sensitive data is information that could cause damage if made public:

- Sensitive personal data; no comprehensive listing of sensitive personal data can be drawn up. The parties conducting the research are responsible for identifying data whose disclosure could harm the study subjects
- Sensitive data may relate to health or the risk of developing a disease, sexual orientation, ethnic background, trade union membership or religious conviction.
- Sensitive species data, such as data concerning endangered animals and plants, data related to nature conservation or biosafety
- Other confidential information, such as patents, trade secrets, military information, or organisational information that could, at the very least, damage you or the organization’s you represent reputation.
- According to its Data Protection Policy (chapter 7), Tampere University is designated as the data controller in research projects that are approved by the University and conducted with core funding granted to the University by the Finnish Government or with external funding provided by an external party.

Personal information includes all identifiers from which the person is identifiable directly or indirectly.

- Direct identifiers: name, phone number, personal identity code, picture, voice, fingerprint, dental chart.
- Indirect identifiers: gender, age, education, profession, nationality, work history, system log history, marital status, residence information, IP address, car license number.
- Remember that spatial data can also contain enough information for identification of individuals.

Further information

- Read more about best practices at Tampere University from Research Data Management Guide.
- Use researchdata@tuni.fi to contact the research data specialists at Tampere University.
- Also, familiarize yourself with the Tampere Universities’ Open Science Policy and Action Plan.
- What is personal data (Finnish Social Science Data Archive)
- Processing of personal data (Office of the Data Protection Ombudsman)

1.2 How will the consistency and quality of data be controlled? (Data Quality)

Explain here how the consistency and quality of data will be controlled and documented when collecting or producing your data. Data quality control ensures that no data is accidentally changed and that the accuracy of data is maintained over its entire life cycle. This may include processes such as repeated measurements, observations or samples, standardised data capture or recording, data entry validation, the peer review of data, or representation with controlled vocabularies. Please note that data quality is separate from the quality of research methods.
Discuss, for example, the following in your answer:

- Which tools you use for collecting or producing your data and how they may affect data quality?
  - What can you do to minimize the risks posed by them?
- How do you ensure that everyone working with your data is familiar with the best practices? For example, how are the interviewers or other data collectors instructed and who is responsible for instructing them?
- Do you use validated methods or quality control pipelines?
- Does someone other than the transcriber check transcriptions of audio or video interviews ("double blinding")?
- Do you use software that produce checksums?
- How do you ensure that the digitalisation of analog or physical material is done with sufficient accuracy?
- How do you ensure that in all conversions the original information content is maintained?
- Who reviews the data before its publication?
- Does your research include processing personal or sensitive data? If yes, Please discuss how minimisation, pseudonymisation, and anonymisation affect data quality.
- Remember that documenting in detail how data are collected or produced provides evidence of data quality.
- AVOID OVERLAPPING WITH THE RESEARCH PLAN! Issues related to data analysis, methods and tools should be described in your research plan - that is, do not include, for example, instrument calibration descriptions here.

2. Ethical and legal compliance

2.1 What legal issues are related to your data management? (For example, GDPR and other legislation affecting in data processing.)

Every research data involves the questions of rights, legal and ethical issues. Show that you are aware of the relevant legislation related to your data processing.

- Check this short video about Legal Aspects of Data by Maria Rehbinder from Aalto University
- Does your data include personal information? If personal data are processed in your research, Data Protection Act will apply to it.
- Does the research include sensitive data (i.e. a specific category of personal data)?
- Does your work require a research permit or ethical review?
- Do you work with other confidential data (e.g., commercial or company data, trade secrets, military information, sensitive species data)?
- Explain what are the risks involved and how are they managed.

Describe here how you will maintain high ethical standards and comply with relevant legislation when managing your research data. In your DMP, the focus should be on the ethical aspects of the data management whereas the ethical issues related to your research methods are covered in the research plan.
Tips for best practices

- Describe how you will ensure privacy protection and data anonymisation or pseudonymisation.
- Specify the party serving as the controller, if you haven’t done that in question 1.1
- Privacy notice helps you plan how you will process personal information during the research process.
- Justify why you have the right to collect, handle, and preserve data that involves ethical issues, for example, that you have passed an ethical review.
- Identify the legal basis for processing personal information (it is usually the public interest or the exercise of public authority.)
- Identify and describe situations when collected personal data are sensitive as well as the legal basis for processing sensitive personal data. If sensitive personal data are processed in the research, the processing must be based on Article 9 of the General Data Protection Regulation (GDPR). Consent from the data subject and scientific purpose of the research are emphasised as a lawful basis when processing sensitive personal data as part of research.
- Pseudonymised data is still personal data, and hence, it must be processed in accordance with GDPR and other data protection legislation.
- AVOID OVERLAPPING WITH THE RESEARCH PLAN! Details of the ethical issues, ethical committee statements and use of laboratory animals should be described in the research plan.

Further information

- Privacy notice
- Tietosuojailmoitus
- Tieteellinen tutkimus ja tietosuoja (in Finnish)
- Informing research participants about the processing of their personal data (FSD)
- General information security guidelines (intra) in Finnish

2.2 How will you manage rights of data you use, produce and share?

Describe how you will agree upon the rights of use related to your research data - both the collected, produced and (re)used data of your project. Here, you can employ your categorisation in the first question. Each of these involves different rights and licenses. Describe the transfer of rights procedures relevant to your project. Describe confidentiality issues if applicable in your project.

- Check this short video about Legal Aspects of Data by Maria Rehbinder from Aalto University
- Who will own copyright and other IPR of the data you will collect, create or reuse?
- How will you license your data for reuse?
- Make an agreement with your group members and state:
  - Who is allowed to use the data during the project
  - What are you planning to do with the data after the project.
  - What license you will use when you open your data
- Are there rights belonging to a third party?
- What kind of data sharing agreements do you plan to make with your research partners?
Agreements on the IPR rights and rights of use of the data should be made before you start collecting or producing the data. By doing so, you prevent possible conflicts at later stages and make sure that your data will be accessible and reusable.

**Tips for best practices**

- Check the [Intellectual property policy](#) of Tampere Universities (chapter 4 and 5). “…The source of funding is the key to determining IP ownership.”
  - “Tampere Universities have an interest in IP arising from **contract research** as this type of IP involves restrictions and third-party rights.”
  - “Employees own IP that they generate while undertaking **open research** and may offer this type of IP to Tampere Universities.”
- Rights to research data may be created in three ways
  1. by legislation (e.g., copyright)
  2. by commitments (e.g., funder requirements)
  3. by agreements (e.g., research consortia, contract research, co-operation with companies).
- If you receive personal or sensitive personal data from a third party, e.g., a health care register, you or the University may still be the controller for that data even if you or the University does not own the data. In these cases, the research contract you make with the data owner sets the conditions how you can use and share the data.
- Data repositories usually requires you to define, under which license you are opening your data.
- [Creative Commons](#) licenses are recommended for research data and metadata. The CC license requires that others who use your work in any way, must give you credit the way you request. [CC0 waiver](#) is the most efficient way of facilitating reuse of your data, but [Creative Commons 4.0](#) is also widely used for open data.
  - [How to select a Creative Commons license?](#)
  - [Data and software licensing wizard](#) by EUDAT
  - [How to license research data](#) by DCC
  - For [open software and code](#), use [MIT, GNU](#) or other appropriate open license.

### 3. Documentation & metadata

3.1 How will you document your data in order to make it findable, accessible, interoperable and re-usable for you and others? What kind of metadata standards, README files or other documentation will you use to help others to understand and use your data? (Metadata & documentation)

- How do you describe the contents of your data? Do you use data dictionaries or [codebooks](#), which explain variables and gather descriptions of codes, calculations and algorithms used?
- Do you have readme file(s) to provide information about data files to ensure they are interpreted correctly?
- What are the [file naming conventions](#) used in your data?
- What kind of directory structure you have?
- Data management software, i.e., databases and an electronic laboratory notebook, which contain details about experiments, analytical methods and the research context.
The purpose of data documentation is to make your data understandable – for you, for your partners, and for the people who will reuse your data. Without proper metadata, your data is useless. Describe the tool, such as Qvain, you will use to document your data sets. Do not mention metadata standards if you do not use them.

Anticipate the opening of metadata or the documentation of your project already here. However, include the detailed description of opening in Section 5 below.

Concentrate here on the documentation of your data sets, that is, the study-level documentation. The data-level documentation and details about experiments, analytical methods and the research context belong to the research plan.

Tips for best practices

- If you use data obtained from a third party it may be accompanied with ready-made documentation.
- Many data repositories require use of a standard metadata format. Hence, if you know where you will store or archive your data, check the requirements for metadata standards.
- It is not essential to know the exact metadata standard when applying for funding. For example, if you are co-operating with a data archive, you can describe that the decision about the standard will be made together with the archive.
- Identify the types of information that should be captured to enable other researchers to discover, access, interpret, use and cite your data.
- Version control. Do you use software of services that automatically creates new version of your files?
- REDCap is a secure web application for building and managing online surveys and databases. The service can be used in Tampere higher education community. Please contact IT-helpdesk for more information.

Further information

- Making a Research Project Understandable - Guide for Data Documentation

4. Storage and backup during the research project

4.1 Where will your data be stored, and how will they be backed up? (Storage & security)

- Which storage services do you use during the research?
- How do you make backup copies in the storage services that you use?
- Do you share the data with your collaborators? If so, how do you ensure the safe transfer of your data.

Describe opening and archiving data after the project in Section 5.

Storage services at Tampere higher education community:
- Personal / Group storage space
  - Protected and backed-up network disk space for members of the Tampere higher education community.
  - For storing sensitive or non-sensitive data (suitable for materials which may not be stored in a cloud service)
- Cloud storage: Office 365 OneDrive for Business (also for external users)
- A virtual server for various purposes (also for sensitive data)
- CSC services for research data
- Do NOT USE external hard drives as the main storing option.
- Please refer to IT for Research for most up-to-date information on the storage services provided by Tampere higher education community.
- If you use commercial cloud services, read their terms and conditions, and makes sure that you are allowed to use them for your purposes.
- Please contact IT Helpdesk 0294 520 500 it-helpdesk@tuni.fi for more information

**Does your research include processing sensitive or confidential data?**

A risk assessment is used to determine the required protective measures for the entire lifecycle of the data.

- Use safe storages for your data (e.g. group or personal storage space, virtual server, ePouta (CSC)
- If your data can be accessed remotely, how is remote access protected?
- Encrypt your mobile devices and external storage devices if needed. If data encryption is necessary please consider following:
  - Which parts of the data are encrypted?
  - Which encryption tools are used?
  - Who manages encryption keys and passwords?
- Do NOT USE external hard drives as the main storing option.
- Do not use cloud storages because they are not safe enough.

**4.2 Who will be responsible for controlling access to your data, and how will secured access be controlled? (Roles & responsibilities)**

Access to personal data must be limited to those individuals for whom it is necessary in order to carry out the research. Please take into account that this group also includes the parties that maintain the services and equipments used, and other external service providers, if any.

- Who is responsible for controlling access to your data?
- How is user and access control implemented?
  - Is there a password protection
  - Do the storage services maintain a log on data use? Remember that you should be able to monitor data usage.
  - Do you have some physical solution (file cabinet) in use?
- Who in the research group has access to which data? Remember that you should have a list of all who have access rights to your data, and a procedure of cancelling rights.
- Why has each access right (editing, watching, deleting) been awarded?
Tell how information security and the risks from sensitive data have been taken into account. Will sensitive data be stored in an encrypted form?

Describe the security of the premises
- How are the facilities used for processing the data protected?
- Can the doors of the facilities be locked?
- Does the property have recording video surveillance?
- Are there burglar-proof storage fixtures/fittings and facilities available for physical material and storage equipment?
- Are the workstations equipped with screen guards?

After they are no longer needed, how will data and copies be disposed in a safe and secure manner?

Tips for best Practices
- With personal and group storage space you can control access rights easily
- Access control should be in line with the level of confidentiality involved.
- Please familiarise yourself with Tampere University’s Information security policy.
- See also how misuse and damage related to personal data are reported.

5. Opening, publishing and archiving the data after the research project

5.1 What part of the data can be made openly available or published? Where and when will the data, or their metadata, be made available? (Data sharing)

Go through all your data types listed in Section 1.1, and answer the following questions:

- What part of the data will be opened or published?
- Where will the data be opened? Name the repositories, if known
- When will the data be available?
- Explain if your data or part of it cannot be opened and give reasons for that. Tell where the project metadata will be opened.

Tampere higher education community’s Open Science Policy defines that the research data produced in Tampere Universities are in principle shared and open. However, confidentiality should not be compromised, and hence, sharing and opening your data should follow the principle: as open as possible, as closed as necessary.

Does your research include processing sensitive or confidential data?

Data with personal information can only be published anonymised, when it is no longer subject to data protection legislation. Pseudonymised data is still personal data and cannot be opened without explicit consent for that purpose.

In some cases, personal information can be shared, if the original processing purpose allows it. However, if the original consent form does not refer to the further use of the data, opening the dataset may require requesting new consent from the data subjects. If you plan to share data which includes personal information, please contact Data protection officer at dpo@tuni.fi.
You should still be able to open the metadata of the data holding personal information, although the actual data cannot be. However, remember that also the file, folder, and variable names may contain personal, sensitive or confidential information.

**Tips for best practices**

- Choose suitable repositories for sharing and opening your data already at the beginning of the project:
  - Check the recommendations of the publishers, learned societies, and funders of your own field of science. Where have you or your colleagues in the same field opened data?
  - List of repositories for different data types can be found in [re3data.org](http://re3data.org).
  - [Zenodo](https://zenodo.org) is easy to use and suitable for any kind of data files smaller than 50gb.
  - Other general purpose repositories are, e.g., [IDA](https://ida.ida.liu.se/), [Dryad](http://datadryad.org), [Figshare](https://figshare.com), [Finnish Social Science Data Archive (FSD)](http://forskningsdatabanken.net), [EUDAT](http://www.eudat.eu) and [GitHub](https://github.com) for sharing code.
  - Check that your data fulfills the repository requirements.
- Metadata, can be opened, for example, in Zenodo, Etsin or forthcoming Tampere University CRIS.
- Check, that the repository assigns persistent identifiers (PID), such as DOI or URN, to your data. A persistent identifier is a long-lasting reference to a digital resource.
- Read more about [opening your research data](http://researchdata@tuni.fi).
- **AVOID OVERLAPPING WITH THE PUBLICATION PLAN!** The research article publication does not equal data publication.

5.2 Where will data with long-term value be archived, and for how long? (Preservation)

The aim of long-term preservation is to keep data usable and comprehensible for tens or even hundreds of years. If your data has long-term value, answer the following questions:

Briefly describe what part of your data you will preserve and for how long. Categorise your data sets according to the anticipated preservation period:

A) Data to be destroyed upon the ending of the project
B) Data to be archived in non-curated archive (e.g. Zenodo) for 20 years
C) Data to be archived by a curated facility for the future generations for tens or hundreds of years.

- Describe the access policy to the archived data.
- Are there some costs related to archiving? Who takes care of them?

Tampere University has determined the process for identifying research data that will retain its value for a longer period and transferring it to Digital Preservation Service for Research Data, [Fairdata-PAS](http://www.fairdata-pas.fo). If you think your data will be suitable for the service, please contact researchdata@tuni.fi.

**Tips for best practices**

- When you start your project and begin to collect or produce your data, consider also how long each data set should be preserved.
Submit your data selected for long-term preservation to a certified data repository or data archive, such as Finnish Social Science Data Archive, Language Bank of Finland or Mendeley Data provided by Elsevier.

- Remember to check publisher, funder, disciplinary or national recommendations for data repositories, data archives or data banks, and their preservation time requirements.

Does your research include processing sensitive or confidential data?

Traditionally, it has been recommended to destroy all sensitive data after the research project has ended, as storing it is risky and requires special arrangements. However, depending on research permits, datasets containing sensitive personal data may also be stored in the Fairdata-PAS service. What is important, is that research participants must be informed about preservation of data and the basis of the duration of preservation.

Archiving datasets that contain sensitive personal data requires a storage permit from the National Archives of Finland. The data must be minimised before storage. The further processing of such data requires a research permit.

- Remember also to plan the safe disposal of the data.
- Please remember that the anonymisation and disposal or archiving of data must be carried out by the expiry of the relevant research permit.
- Genuine anonymisation requires that both direct and indirect identification are made impossible, in addition to which the identification key must be destroyed.
- Check the five steps in deciding what data to keep (DCC, UK)
- Use researchdata@tuni.fi to contact the research data specialists at Tampere University.

6. Data management responsibilities and resources

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

Summarize here all roles and responsibilities you have described in previous answers.

Tips for best practices

- Outline the roles and responsibilities for data management/stewardship activities for example data capture, metadata production, data quality, storage and backup, data archiving, and data sharing. Name responsible individual(s) where possible.
- For collaborative projects, explain the co-ordination of data management responsibilities across partners.
- Indicate who is responsible for implementing the DMP, and for ensuring it is reviewed and, if necessary, revised.
- Consider regular updates of the DMP.
- Also, consider who will be responsible for the data resulting from your project after your project has ended.
Does your research include processing sensitive or confidential data?

- Who is responsible for the management of sensitive and confidential data as well as monitoring its implementation throughout the lifecycle of the data?
- Who is responsible for data protection (see Section 2) and information security (see Section 4)?

Tips for best practices

- Specify your data archiving, opening, and publishing costs in the budget.
- How will responsibilities for data management and costs be split across partner sites in collaborative research projects?

6.2 What resources will be required by your data management to ensure that data can be opened and preserved according the FAIR principles (Findable, Accessible, Interoperable, Re-usable)?

- What resources (time and workload) are needed for data management?
  - Consider if additional computational facilities and resources need to be accessed, and what will be the costs associated with this.
- Give an estimate of how much time is needed for data documentation and cleaning (i.e., producing metadata, anonymising sensitive data, arranging data, transferring data etc.) to prepare the data (not results) to be opened: 1–2 h weekly, one day per month, 1–2 weeks before publishing, or some other time estimate.
- When planning the resources needed, take into account requirements set by a higher level of security for activities and solutions used, as well as related additional costs.
  - Costs of data minimisation, pseudonymisation and anonymisation, or the time spent and software required for the task

Putting data into a usable format and making it meaningful to other researchers takes time and costs money in terms of software, hardware, and personnel. Start planning your data management already in the beginning of the project, and you will save time and effort when you are opening and preserving your data.