UNIVERSITEIT TWENTE.

Data Management Plan (info)

General information (info)

Title of the research project	
Name of student/researcher(s)	
Name of supervisors/(co-) promotors	
Name of research group	
Description of your research	*Briefly summarise the type of your research to help others understand the purposes for which the data are being collected or created.*
Funding body(ies)	
Grant number	*A grant number provides unique identification for the grant.*
Partner organisations	
Project duration	Start: **MM-DD-YYYY** End: **MM-DD-YYYY**
Date written	**MM-DD-YYYY**
Date last update	
Version	*A new version of the DMP should be created whenever important changes to the project occur due to inclusion of new data sets, changes in consortium policies or external factors.*
Name of researcher(s), the institution/group and their roles/responsibilities for data management	*Name anyone and their specific roles and responsibilities for data management (data collection, data analysis, data storage, etc. as well as data ownership).*
Laws, policies, contracts and agreements to comply with	*When applicable, name the laws, data policy documents and contracts or agreements to which this DMP must comply.*

1. Data Collection (info)

Describe the data you will collect. The answers on most of these questions can be placed in the table below.

1. What type of data will be collected or generated? (measurements, observations, questionnaires, models, etc.)
2. Will you also use pre-existing data (secondary data)? From which source?
3. How will data be collected or generated?
4. What file format the different types of data will have?
5. Which tools or software are needed to create, process and/or visualize the data?
6. What is the estimated total volume of the data to be managed and/or used? What is the estimated number of files and the maximum file size?
7. Are there rights and responsibilities which need special attention regarding the use and management of the data (especially in case of secondary data)? Which (legal) agreements, consents, assessments or licences are relevant or should be made regarding data collection, processing or (re)use?
8. Are there third parties claiming the (shared) copyright on data collected, generated or used in your research? Name these parties and refer to the agreement in which copyright on the data is arranged.
9. Do you collect or use personal data in terms of the European General Data Protection Regulation (GDPR)? According to the *Code of Conduct for the use of personal in academic research* is it necessary to report this to the Dutch Data Protection Authority? (See guidance below).
10. What are the estimated costs involved in collecting, generating or acquiring/using data?

Type of data (1.1)	Format (1.4)	Software (1.5)	Total data volume (1.6)	Number of files (1.6)	Personal data (1.9)	Costs (1.10)

2. Data Documentation (info)

Document your data to help others to understand and reuse it.

1. What standards will be used for documentation and metadata? If there is not a standard already available for your data, outline how and what metadata will be created.
2. How will your data be documented during your research and for long-term preservation?
3. What directory and file naming convention will be used to enable the titling of your folders, documents and records in a consistent and logical way?
4. How do you handle version control to maintain all changes that are made to the data?
5. What are the estimated costs involved in documenting the data?

3. Data Storage (info)

Ensure that during your research all research data are stored securely and backed up or copied regularly. The answers on these questions can be placed in the table below.

1. Specify what data (raw data, processed data, non-digital data, etc) and related materials (models/codes, informed consents, etc.) will be stored and backed up during the research?
2. Which storage medium will you use for the master files of your data?
3. In case of non-UT master file storage, what is the backup procedure?
4. Which storage medium will you use for copies of your data files?
5. In case of non-UT storage media, what will be your backup strategy in terms of location(s) and frequency?
6. What are the estimated costs involved in storing the data and related materials?

Data (3.1)	**Storage medium (3.2)**	**Backup frequency (3.5)**	**Backup location (3.5)**	**Costs (3.6)**

4. Data Security (info)

Describe how access to the data is managed and authorized both during and after the research.

1. Do the data you collect, generate and/or use have a confidential character, and if so, for what reasons?
2. In case of confidential data, is it possible and sensible to split up your data in confidential and non-confidential sets?
3. During the research, who controls access to the data (e.g. researcher, Principal Investigator, student, lab, university, funder)?
4. During the research, what security measures will be in place to prevent unauthorized access to the confidential data? In case of data encryption, how this will be implemented?
5. During the research, apart from you as the researcher, who is allowed to have access to the confidential data and how do you arrange this securely?
6. After the research, who controls access to the data (e.g. researcher, Principal Investigator, student, lab, university, funder)?
7. After the research, what security measures will be in place to prevent unauthorized access to the confidential data?
8. After the research, apart from you as the researcher, who is allowed to have access to the confidential data and how do you arrange this securely?
9. What are the estimated costs involved in data security?

5. Data Selection and Preservation (info)

Describe which data will be selected for preservation after the research and where and how these will be archived. The answers on 5.1, 5.3 and 5.5 can be placed in the table below.

1. Which data and related materials should be archived for preservation and long-term accessibility. Which data and related materials can or has to be destroyed?
2. Are there data of others to be preserved? If so, how this will be arranged in such a way that these data are preserved for verification and/or reproducibility?
3. What file formats will be used for long-term preservation and accessibility?
4. What is the estimated total volume of the selected data to be preserved? What is the estimated number of files and the maximum file size?
5. After the end date of the project, how many years the selected data should be preserved?
6. What are the estimated costs involved in the preparation of the data for preservation?

Data / related materials (5.1)	Preservation format (5.3)	Preservation period (5.5)	Preparation costs (5.6)	Data archive (6.1)	Publicly available (6.2 / 6.3)

6. Data Availability for Reuse (info)

Describe how your data can be made available and reused. The answers 6.1, 6.2 and 6.3 can be placed in the table above.

1. Which repository is appropriate for archiving the selected data (4TU.ResearchData, DANS, subject-based data repository)?
2. Which of the selected data for preservation can or must be made publicly available (for reuse)?
3. Which of the selected data for preservation cannot be made publicly available and for what reasons (e.g. legal restrictions on personal data, intellectual property, commercial interests, security)? Could the data be split up into publicly and non-publicly available sets, and if so, how?
4. What are the estimated total costs (€) for archiving the data and related materials?
5. Which policies or agreements are relevant for data sharing apart from the UT research data policy (e.g. funder agreements, other university policies, third parties agreement)?
6. If you allow others to reuse your data, how will the data be shared (e.g. general or subject data repository, website, on request)?
7. When making the data available for reuse, are you going to offer the standard license for data availability (CC0 or CC-BY) or another license? Which one?
8. Are you considering to publish your data in a data paper? If so, in which data journal?

Guidance

Data Management Plan

This template should be regarded as a checklist of most important questions when formulating a data management plan. Be aware of specific issues which might be relevant in your research but are not or only briefly touched in this template.
The template may be appropriate for Principal Investigators collaborating in a research project or for research students working on a PhD or Masters project.
The Data Management Plan consists of a page for general information and 6 sections.
Detailed information on each section will show up when you 'open' the info button.

General information

Apart from general information about the research, researcher, partners, funding agency, etc. pay attention to the question about roles and responsibilities. This can refer to the different parts of the DMP, like data collection or data access, but also to checking the *quality* of data and data documentation. Also think about the roles and responsibilities of monitoring the DMP implementation and the DMP compliance with laws, policies, etc. Ask whether there is information about this in your group, institute or faculty.

Make distinction between who should perform a certain action and who checks the results of these actions. Moreover keep in mind the distinction between data management actions during and after the project.

Have a closer look at the copyrights of the data, or in most cases more important, the exploitation rights. There might be policies (the UT research data management policy in the first place), contracts or agreements about this. If not, discuss this issue with involved persons.

Mention all official laws, policies, contracts and/or agreements which are relevant to the research and say something about the data to be collected and/or used in the research project and how these should be handled.

1. Data Collection

Research data is information in any form collected or generated during research. It is used as a basis for analyses. The data creation phase starts with deciding what kind of data you need to answer your research question(s). If you intend to reuse existing data instead or additional to collecting research data yourself, there are good sources for finding potentially relevant existing data.

Give a summary of the data you will collect or create, indicating the content, coverage volume, data formats and data type, e.g., text, numbers, tabular data, survey data, experimental measurements, models, software, audio-visual data, physical samples, etc., as these have an impact on the way you can process, store, share, preserve and access your data. Providing a table with characteristics of your data might be a convenient way to answer this question.

Data types

There are four main types of research data:

- **Observational data**: captured in real time, typically cannot be reproduced exactly
- **Experimental data**: from labs and equipment, can often be reproduced but may be expensive to do so
- **Simulation data**: from models, can typically be reproduced if the input data is known
- **Derived or compiled data**: after data mining or statistical analysis has been done, can be reproduced if analysis is documented

Data types can include text, numbers, images, 3D models, software, audio files, video files, reports, surveys, etc.

Also take into account related material (used for or during collection, processing and/or analysis) which should be managed together with the data like scripts, questionnaire forms, informed consents, etc.

Provide information on the existence (or not) of similar data and the possibilities for integration and reuse.

File formats
In planning a research project, it is important that you consider which file formats you will use to store your data. In some cases, this will be dictated by the software you are using or the conventions of your discipline, but in other cases you may have to make a choice between several options. These are likely to be some of the key factors in your decision-making:

- what software and formats you or colleagues have used in past projects,
- any discipline-specific norms (and any peer support that comes with them),
- what software is compatible with hardware you already have,
- whether you have funding for new software for the job,
- how you plan to analyse, sort, or store your data.

But you should also consider:

- what formats will be easiest to share with colleagues for future projects,
- what formats are at risk of obsolescence, because of new versions or their dependence on particular software and/or hardware,
- what formats it will be possible to open and read in the future,
- what formats will be easiest to annotate with metadata so that you and others can interpret them days, months, or years in the future.

In some cases you may be best off using one format for data collection and analysis and converting your data to a standard format for archiving once your project is complete. After conversions, data should be checked for errors or changes that may be caused by the export process.

How data will be collected
You can briefly describe here the methodologies you will use to collect or generate your research data. These may include interviews, observations, machine measurements, etc. If possible specify which measurements you will perform, e.g. MRI-scans, GPS-tracking, etc. Methods can determine how accurate your results will be, and how interoperable with other research.

You can briefly explain:

- How the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeat samples of measurements, standardised data capture, data entry validation, peer review of data or representation with controlled vocabularies;
- How your data could complement and integrate with existing data, or whether there are standard methods that you use.

Personal data
In case you collect personal data, data on identified or identifiable natural living persons, explain what actions you will pursue to safeguard the privacy of persons involved in your research or to protect the confidentiality of your data. Make sure to consider transparency, data minimisation and purpose, storage, and access limitation as appropriate measures.

In the European General Data Protection Regulation for privacy sensitive data, accuracy of data is a guiding principle. Check the UT website for more information.

There are different protection regimes, dependent on the type of personal data. There is stricter regime for ***special personal data***, like sexual orientation, religion, criminal record, political affiliation. Even more strict protection is required for patient- and BSN (citizen service number) -data. See for more information: *Code of Conduct for the use of personal data in academic research*.

Before the start of collecting personal data there is basically always informed consent needed for storing, processing or publishing.

Rights issues
In the first place you will have to distinguish whether you collect or generate the research data by yourself or in the framework of your own research (primary data) or you use data collected by another, outside the framework of your research (secondary data).

In case of primary data, intellectual property rights ("database right") is vested in the University of Twente (see UT research data policy, section 4). Shared copyright with a third party should be arranged in a specific agreement. Make clear arrangements about the responsibilities regarding the use, sharing and reuse of the data, both during and after the project.

In case of secondary data, often you can use these under confidentiality or other restrictions. You must identify and explain how you will deal with these restrictions in your data management plan, especially who you will deal with availability of the data for verification and reuse.

Costs
There may be costs involved in collecting, generating or acquiring/using data. Use this guide for estimating these costs.

2. Data Documentation

Describe the types of documentation that will accompany the data to help secondary users to understand and reuse it. This should at least include basic details that will help people to find the data, including who created or contributed to the data, its title, date of creation and under what conditions it can be accessed.

Documentation may also include details on the methodology used, analytical and procedural information, definitions of variables, vocabularies, units of measurement, any assumptions made, and the format and file type of the data. Consider how you will capture this information and where it will be recorded. Wherever possible you should identify and use existing community standards. See: http://www.dcc.ac.uk/resources/metadata-standards

File naming
Organising your files and folders effectively and efficiently can save you time and make collaboration easier by ensuring you are working on the correct version of the data. A good file name makes it easy to identify, locate and retrieve your data. There is no one recommended way to name your files and folders, but you should name your files consistently. If you work as part of a research group, you should decide on a file and folding naming system with your colleagues. See for practical information: http://guides.lib.purdue.edu/c.php?g=353013&p=2378293

Identifiers
An identifier is a reference number or name for a data object and forms a key part of your documentation and metadata. To be useful over the long-term, identifiers need to be unique (globally unique if possible) and persistent (the identifier should not change over time).

The emerging identifier standard for publicly available datasets is the Digital Object Identifiers (DOIs). Although DOIs have been traditionally used for journal articles, they can

now be assigned to datasets. 4TU.ResearchData (more information see section 6: Data preservation and archiving) will automatically assign a DOI to a dataset that you make available.

Version control
Because digital research data can so easily be copied, over-written or changed, researchers need to take steps to protect its authenticity. Research time is wasted and valuable data put at risk if researchers work with outdated versions of files.
Version control can prevent this. Control is particularly important if data is being used by multiple members of a research team, or if research files are shared across different locations.

A regime to synchronize different copies or versions of files will improve research efficiency and help guarantee the authenticity of the data. Good practice generally involves the keeping of a single master file, to which all changes are recorded. Version control mechanisms should be established and documented before any data is collected or generated.

Read more about version control at: http://www2.le.ac.uk/services/research-data/organise-data/version-control
There are open software tools for version control. See:
http://www.unmc.edu/vcr/rito/services/version-control-handout.pdf

Costs
There may be costs involved in documenting the data. Use this guide for more information.

3. Data Storage

It is the responsibility of the researcher to ensure that their research data and related information like scripts, software, survey templates, documentation, informed consents, etc. is stored securely regularly backed-up during the life of the project. It is UT policy to store the master files of the research data on one of the university's networked fileservers. Copies can be kept on remote, cloud and/or portable storage, depending on confidentiality requirements.

Generally there are four options for data storage:

- **University fileservers: Home Directory and Project or organization directory** – As these are secure and backed-up regularly, these storage media are ideal for storing the master files of your research data.
- **Local drives: PCs and Laptops** – Data can be lost because local drives can fail, or the computer may be lost or stolen. These are convenient for short-term storage and data processing but should not be relied upon for storing master files.
- **Remote or Cloud storage** – For cloud storage it is recommended to use Surfdrive, also for confidential data. Commonly used services, such as Dropbox and Google Drive on a personal account, is not appropriate for sensitive data. For non-confidential data these cloud services should only be used for copies. However, cloud services as Google Drive and OneCloud *accessed by a UT account* can be used, even for confidential data.
- **External portable storage devices** – External hard drives, USB drives, DVDs and CDs are not recommended for storage as their longevity is uncertain and they can be easily lost or damaged.

For more information about storing your research data, see:
https://www.utwente.nl/en/lisa/researchsupport/perform/#storing-and-sharing-research-data

You may choose to only back up certain data, or to back up files you use every day more regularly than others. The basic rule of thumb is: The more important the data and the more often they change, the more regularly they need to be backed up.

If your files take up a large amount of space and backing up all of them (or backing them up sufficiently frequently) would be difficult or expensive, you may want to focus on backing up specific key information, programs, algorithms, or documentations that you would need in order to re-create the data in case of data loss.

For more information about backups, RPO and RTO, see:
https://www.utwente.nl/en/cfm/services-abc/!/product/p883320/backups-for-employees

Costs
There may be costs involved in storing the data. Use this guide for more information.

4. Data security

Check the authorization of access to the data with legal regulations and contracts related to your research. Especially when third parties are involved and in case of confidential research data, copyrights and responsibilities regarding access, use and sharing/reuse must be clarified prior to, or at the beginning of a project.

(Part of) the research data you collect, generate or uses may have a confidential character. In most cases this has a permanent status, which means that these data must be handled in accordance with legal or contractual regulations both during and after the research. You need to take data security measures. By classifying your data you can determine which measures should be taken, who has access to your data and what they are authorised to do with it. Consider the option of splitting up your data in confidential and non-confidential sets.

Data security is needed to prevent unauthorised access or disclosure and changes to or destruction of data. The principle investigators are responsible for ensuring data security. The level of security required depends upon the nature of the data – personal or sensitive data need higher levels of security (see above: Data collection – personal data).

It is possible that you will need remote access to your data, if you are working from more than one location, or not at the university. A number of individuals may require access to the data, possibly with different privileges to read, write, update or delete. This may be accomplished by keeping a copy of the data on the university shared network file store, where it is password protected. The use of cloud storage to share data depends upon the level of security needed.

It is possible that your project may need to arrange for access to third party data that may have specific limitations in how they can be distributed (based on IP or the agreement by which your project obtained the data). When your research project has received data under confidentiality or other restrictions you will have to identify and explain these restrictions in your data management plan.

Costs
There may be costs involved in data security. Use this guide for more information.

5. Data Selection and Preservation

You need to decide, together with the project stakeholders, which data have to be preserved and long-term accessible after the end of the research. The same holds for data related materials like scripts, software, survey templates, documentation, informed consents, etc. The decision will be based on the standards of good scientific practice, legal and contractual regulations, funder requirements, but also on the type of data created, the value for reuse, and whether further work or publications will be based on it.

Data selected for long-term preservation will normally be submitted to a funder established data centre, disciplinary data repository or an institutional data repository. In the Netherlands you can choose 4TU.ResearchData for the technical-scientific research data and

DANS for data from research in the social sciences and humanities.

4TU.ResearchData stores the data in a permanent and sustainable manner, according to the guidelines of the international Data Seal of Approval. Being a Trusted Digital Repository, 4TU.ResearchData is demonstrating to researchers that it is taking appropriate measures to ensure the long-term accessibility and quality of data it holds. If you need any help with depositing your data please contact 4TU.ResearchData for assistance.

The conditions under which the data may be made available by the data repository to other researchers are determined by you as the Principal Investigator depositing the data. When depositing to 4TU.ResearchData you can choose from two general access conditions:

- **Open access**: there are no additional restrictions on access to the data or publication of results.
- **Embargo period**: you can request that an embargo period be imposed on your data, whereby no access to the data would be permitted until after the date you specify.

At DANS (Data Archiving and Networked Services) research data are sustainably stored (it also has the Data Seal of Approval) and shared via EASY, the DANS online archiving system. For help and more information you can contact info@dans.knaw.nl

Other data repositories can be found at Registry of Research Data Repository.

Costs
There may be costs involved in the preparation of data for archiving. Use this guide for more information.

6. Data Availability for Reuse

Your research is valuable and important, and so is the data that it is based on. By sharing your data, you make it available to the scholarly community, who can study and build upon your work. Your work will become more visible and typically be cited more frequently.

Think about how you want others to reuse your data. If you want your data to be as widely used as possible, the Creative Commons Zero (CC0) or the Attribution Only licence (CC-BY), would be most useful. These licenses let others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation (CC0 not according to the license, but according to conduct code of scientific behaviour).

Some data repositories have licenses that depositors must grant as a condition of deposit. When depositing to 4TU.ResearchData you must sign a licence agreement to establish the terms and conditions of use of your data collection. This is a legal document which sets out your rights and responsibilities as depositor and ours as the data distributor.

Depositors can elect to apply an embargo to the research data so that public access is deferred for a specific period (typically no more than two years). Embargo may be appropriate when the researcher needs to maintain the data in a managed repository environment, like 4TU.ResearchData, while deferring any access to the data pending further data collection, analysis, publication of results, etc. If data are generated using specifically developed software, it may be necessary to provide a copy of the software, noting operating requirements, with the data.

At the end of your research project, your funder may require you to share your research data, by publishing it with no access restrictions (open access). Some journal publishers also require the data supporting the research article to be published.

Publishing by means of a paper in a data journal may be interesting for your visibility and scientific reputation. There are specific data journals in which you can publish a data paper. Find more information at Sources of Dataset Peer Review of the University of Edinburgh.

Costs
There may be costs involved in sharing and/or archiving the data. Use this guide for more information.