**Data Management Plan template @ Utrecht University**

*(May 2018)*

Plan details:

* The data management plan is a living document, you can update this as your project takes shape.
* We recommend to use the standard Utrecht University guidance which is available at [DMPonline](https://www.uu.nl/en/research/research-data-management/tools-services/tool-to-create-your-dmp-online) when you select the UU template.
* Utrecht University offers an online self-study training ‘[Learn to write your DMP’](https://lll-platform-uu.nl/login/index.php)
* Example answers can be used as possible answers by researchers.

**PREPARE**

1. Data collection
2. Data documentation

# Data collection

Consider all the research data you will be collecting for your research, including any existing data or third-party sources that will be used.

**Please answer the following questions**:

1.1 Will you use existing data?

1.2 What data will you collect or create?

1.3 How will the data be collected or created?

1.4 How will you manage rights issues?

1.5 What are the costs involved in managing and storing data

**More information**

If you want more information on data collection you can check the [RDM website of Utrecht University](https://www.uu.nl/en/research/research-data-management).

## Will you use existing data?

* No
Please specify why existing data is not sufficient for your research.
* Yes, we will use data from:
* Previous research (own/group);
* An academic collaboration;
* Commercial collaborators;
* A publicly available database/archive;
* A specialist (commercial) data provider;
* Another data provider (please specify).

**UU Guidance:**
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.

**Example answer**

No. Preliminary literature review showed that although a great deal of data exists about the technological aspects, there is little data available which pertains to the social and governance aspects. There is thus a need to gather primary data on this topic to answer our research questions.

**Example answer**

YES - We will use data from CBS (another data provider).

## What data will you collect or create?

Give a brief description of the data, including the type, volume (if known) format and content. If possible, give a rough estimate of the number of files.

**UU Guidance on Data Description**

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.

Consider the following:

* Do you have digital or physical data?
* What type of data will be collected or created?
* What is the expected volume of your research data?
* What data formats will you be using?
* What kind of tools or software do you need to create, process or visualise the data?

**UU Guidance on Data Volume**Indicate the expected volume of the data you will create. Note the volume in MB/GB/TB/PB. Google may be of help in finding examples of file volumes.

Consider the proportions of

* raw data (unprocessed, as you obtained them),
* processed data (transformed to make further analyses possible),
* other, secondary outputs (e.g., reports).

**UU Guidance on Data Format**Indicate which file formats you will use for storing your data. Examples of file format(s) are plain text (.txt), comma-separated values (.csv), geo-referenced TIFF (.tif, .tfw). Explain why you have chosen certain formats. Your decision may be based on the software you use, the conventions of your discipline, staff expertise, a preference for open formats, the standards accepted by data centres or widespread usage within a given community.

. Consider the following: Not all formats are suitable for long-term retention or exchanging data with others. Using standardised, interchangeable or open formats ensures the long-term usability of data. That is why these are recommended for sharing and archiving.

In some cases, you may be best off using one format for data collection and processing and converting your data to a standard format for analysis or upon archiving once your project is complete. Note: After conversion your data should be checked for errors or changes.

**Example answer**

We will have audio files from the interaction experiment. These will be transcribed to text files. We will have one audio/text file per experiment (23 in total). The audio files will be 250 MB each, the text files will be 1 MB each. In total this amounts to almost 6 GB. Formats for audio will be standard .wmv format, and transcriptions will be .txt format. These formats are commonly used and can be opened by many (freely available) software programs.

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Description** | **Files (number)** | **Total volume** | **Format** |
| Audio files from the interaction experiment. | 23 | 5750 MB | .wmv |
| Transcriptions of audio files | 23 | 23 MB | .txt |

## How will the data be collected or created?

Briefly describe the research methodologies used and how you will ensure data quality.

**UU Guidance on Data Collection**

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.

In the European General Data Protection Regulation for privacy sensitive data, accuracy of data is a guiding principle.

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.

**Example answer**

We will collect data on the sexual and religious preferences of Dutch male participants between 18 – 20 years using LimeSurvey, an online questionnaire application. The questions will be standardised between the participants, and where possible also the answer possibilities with multiple choice. We found one dataset in the UK Data Archive that aligns to our study. To make results interoperable with the identified existing questionnaire data as found in the UK Data Archive, we will align questions and answer options to that existing source where possible. In transferring the results to a single result sheet, a script to read the individual result files and add them to the sheet is used to ensure quality of transfer. Two random files will be manually checked for consistency in the sheet.

## How will you manage rights issues?

State who will own the copyright and intellectual property rights (IPR) on any data that you will collect or create, along with the agreements you made for its use and reuse.

**UU guidance on Intellectual Property Rights**

When your research project has received data under confidentiality or other restrictions, you will have to identify and explain how you will deal with these restrictions in your data management plan.

Make sure that you:

* State who will own the copyright and IPR of any existing data as well as new data that you will generate;
* For multi-partner projects, IPR ownership should be covered in the consortium agreement;
* Outline any restrictions needed on data sharing, e.g., to protect proprietary or patentable data;
* Consider negotiating restrictions on the reuse (and subsequent sharing) of any third-party data you will use.
* Note that access by others to external data will also be necessary to enable others to verify results from your research.

RDM Support offers you the guide [Legal instruments and agreements](https://www.uu.nl/en/research/research-data-management/guides/legal-instruments-and-agreements). This is an overview of legal agreements you can make with different parties, either people involved in your research, third parties or data users.

**Example answer**1) We will use sources from private archives. These sources are not allowed to leave the archive, also not in digital form. Extracts of the sources are allowed to leave the archive, and we will use these to give an indication of the content of the original sources.
2) Some external sources we want to use are licenced under a CC BY NC SA licence, others under CC BY SA. As we want to combine these, but these licences are incompatible, we will negotiate with the authors if our new combined product can be released under a simple and permissive CC BY licence, while also referencing the original sources, to avoid such problems in the future.

## What are the costs involved in managing and storing your data

There may be costs involved in managing and storing your data. Specify these costs here.

**UU guidance**

You can specify any costs here for data management that you identify.

On the RDM website of Utrecht you’ll find the guide to [Costs of Data Management](https://www.uu.nl/en/research/research-data-management/guides/costs-of-data-management). Please read the items and description and write down

* if they are relevant for you and if so,
* how the costs are covered.

Costs can be covered in the overhead of your university, faculty, or be reimbursed by your funder if you budgeted for this in your application.

**Example answer**

We will acquire a licence to use the commercially available dataset on household statistics, which will amount to 6.000 euro for three years. Costs are covered in the funding budget for materials. For data cleaning and preparing for analyses, we will hire a programmer for a total of 2.000 euro. The data that we ourselves will collect and generate, we will store in YODA, for 16 euro per TB/month, we will also use the YODA solution for long term archiving (10 years). These costs are covered by the Faculty. In addition the data are published for long term accessibility in DRYAD repository, for 120 dollars total (max. 20 GB). These costs are covered by the overhead budget in the funding.

# Data documentation

Proper description and documentation of your data will make your research process more transparent and reproducible. It will enable reuse, by yourself and others.

**Please answer the following questions**:

2.1 How will you structure your data?

2.2 How do you handle version control?

2.3 How will the data be described and documented?

**More information**

If you want more information on data documentation you can check the [RDM website of Utrecht University](https://www.uu.nl/en/research/research-data-management/guides/storing-and-preserving-data/data-description-in-practice).

## How will you structure your data?

Describe your intended folder structure, and file naming conventions.

**UU Guidance**
Indicate how your data will be organised during the project, mentioning, e.g., naming conventions, and folder structures. If you work as part of a research group, you can decide on a file and folder naming system with your colleagues.

Consider that:

* Organising your files and folders effectively and efficiently can save you time.
* Consistent, well-ordered research data will be easier to find, understand and reuse.

**Example answer**

For file names, we will use the coded elements. ‘ma’ for microarray data, ‘pt’ for protein data, ‘wp1’ for work packages that generated the data, ‘gr1’ for the measurement groups and ‘180921’ for dates of measurements. All coded elements are described in a separate document X. The complete folder structure is outlined in a separate diagram Y, an extract is given here:

>Project folder

 >>Results folder

 >>>Exp1

 >>>Exp2

 >>Outreach folder

 >>>Public\_Engagement

## How do you handle version control?

Describe how you will keep track of changes to your research data.

**UU Guidance**

Because digital research data can so easily be copied, over-written or changed, it can be useful to take steps to protect its authenticity.

Version control will allow you to keep track of any changes you make to your research data and to prevent a mix-up of old and new versions.

Here are some options for version control to consider:

* Use version control software (if you use this, please specify which software);
* Include the date/version number in filename/folder;
* Use the ‘Track changes’ feature in software;
* Process your data with a script, save the script with which you process your data;
* Delete all minor versions at set times.

**Example answer**

We will distinguish versions by indicating the version in the filename of the master copy by adding a code after each edit, for example V1.1 (first number for major versions with a significant change, last for minor versions which could be text edits). The most recent copy at the master location is always used as the source, and before any editing, this file is saved with the new version code in the filename. The file with the highest code number is the most recent version. Every month, we will delete minor versions. The major versions will be listed in a version document (projxVersDoc.txt), stating the distinguishing elements per listed version.

## How will the data be described and documented?

Briefly describe how peers should be able to understand your data.

**UU Guidance Metadata & Documentation**

Documentation and metadata are needed to accompany your data to help yourself, co-workers, or secondary users verify, understand and/or reuse the data. Documentation is anything human readable that describes the data, metadata are (standardized) fields that take a value (e.g. species: rat or air pressure: 800) to describe characteristics or specific circumstances of samples you took. Documentating and filling in values to metadata fields for your data is preferably done during the data generation, as it is often impossible or very time-consuming to assign afterwards.

Consider:

* Indicating what, where and when you will have the necessary documentation to make sure your research can be reproduced and understood. Documentation can include lab journals, a codebook explaining your variable names, survey questions or exact scripts for calculations used with analysis of the data, parameter settings, detailed methodology; etc, etc, Indicate if there is a standard for documentation in your field;
* Stating what metadata you will provide on study level to help others identify and discover the data as a set. Note that when using Yoda (for i-Lab, for Youth or for a specific faculty), standard metadata on study level will also be added;
* Using community metadata standards where these are in place. The Research Data Alliance offers a [Directory of Metadata Standards](http://rd-alliance.github.io/metadata-directory/). Data repositories may also provide guidance about appropriate metadata standards;
* How you will capture metadata and data documentation and where it will be recorded, e.g., in a database with links to each item, in a ‘read me’ text file, in file headers, the associated publication, etc.

**Example answer**

We will give a description of the context of how the samples were generated in file X. The measurements on the samples will be according to the method of standard weighing, which will be referred to. All parameter settings will be kept as a separate file together with the measurement results. A spreadsheet will be designed which will hold a list of all experimental information which is mandatory for all collaborating partners to note down. A codebook will explain the meaning of the column names in the spreadsheet. A script with the code to process and analyse the data files will be saved as a file, together with the results. Coded elements in the file names as described in 2.1 will indicate the cross-references between scripts and results.

**HANDLE**

1. Data storage
2. Data security

# 3. Data storage

How do you ensure that during your research all your research data is stored securely and backed up regularly?

**Please answer the following questions**:

3.1. Where will you store your data?

3.2. How will the data be backed up?

**More information**

If you want more information on data storage you can check the IT solutions on the [RDM website of Utrecht University](https://www.uu.nl/en/research/research-data-management/tools-services/tools-for-storing-and-managing-data).

**3.1. Where will you store your data?**

List the locations where your data will be stored, and indicate if storage capacity is sufficient.

**UU Guidance on storage and security**

A good overview of where you store all your data will avoid loss of data or a mix-up of data files. For each storage location check if it meets your needs. Does it provide the required capacity? Can you work on the data with others?

Consider:

* Your storage locations may vary depending on your research activity (e.g., doing fieldwork or working across multiple sites involves different solutions);
* It is efficient if every data file has one single location for the official and most recent version, this is called your master copy location;
* Identify at which other (temporary) locations you or your collaborators will store your data.
* Describe if you have a separate location to store your raw data during research. In general, raw data should be stored as early as possible, preferably in a write-protected ‘vault’, time-stamped if possible, so it will remain unaltered at all times;
* Describe where any non-digital data or outputs that the project will generate will be stored (if applicable).

Consult [RDM Support](https://www.uu.nl/en/research/research-data-management/guides/storing-and-preserving-data/data-description-in-practice) for advice about secure storage options and other infrastructure solutions specifically designed to store and backup (high volumes of) research data.

Note that (privacy-)sensitive data needs additional measures for storage (see 4. Data security).

**Example answer**

The data of work package 1 is stored on Utrecht Universities O:drive (maximum capacity 4GB). For work package 2, cooperation with University of Toronto is required. Therefore we will store the data for that work package on the cloud service SURFdrive (maximum capacity 250 GB) and grant access to the partners in Toronto. All original raw data is stored in a separate, write-protected folder on the O: drive and analyses are done on working copies. Regularly we will do a check if the working copies remain the same as the originals, by comparing their checksum. The data stored on SURFdrive will be downloaded for processing by various members of our team, we will have the most recent version stored there every day before 8 pm.

**3.2 How will the data be backed up?**

Describe your backup strategy or the automated backup strategy of your storage locations.

**UU Guidance**

You are responsible to ensure that your research data is regularly backed up and stored securely. Make sure that you:

* Indicate if storage locations have automated backup such as is the case in YODA, University networked drives, SurfDrive, and if not, describe how you yourself will back up your data at different locations. It is good practice to keep at least three copies of crucial data, one at a remote location;
* Describe how your data will be recovered in the event of an incident. Indicate how you will make backups with sufficient frequency to be able to recover your data in the event of data loss;
* Describe who will be responsible for backup and recovery.

**Example answer:**

1 From my master copy location, I will back up my changed data every day to my hard drive, and once a week to a drive on another location. In this way, I will at most lose a few hours of work, or a week in more rare cases if disaster strikes my master copy location. Every month I will do a full backup and keep that for half a year. Once a year I will do a full backup and keep it for a year. I have checked the accessibility of my backups and I’ve made sure that I can restore my files if necessary.

2 I will store my data at the University Networked drives, from which twice a day a backup is made, that within a few hours is also transferred to a separate location. Every month and year backups are made for prolonged storage.

# 4. Data security

One of the aims of good data management is to safeguard your data from unwanted access and misuse and prevent them from being lost. When you are collecting confidential or privacy-sensitive data, it is even more important to think about security issues.

**Please answer the following questions**:

4.1. Will you use or collect any confidential or privacy-sensitive data?

4.2. How will you handle confidential or privacy-sensitive data?

4.3. What measures will you take to comply with security requirements and mitigate risks?

4.4. To whom will access be granted/restricted?

**More information**

If you want more information on data security you can check the RDM website of Utrecht University on [handling personal data](https://www.uu.nl/en/research/research-data-management/guides/handling-personal-data) or on [storing and preserving data](https://www.uu.nl/en/research/research-data-management/guides/storing-and-preserving-data).

**4.1. Will you use any confidential or privacy-sensitive data?**

* Yes > Please continue to question4.2 and omit 4.3 to continue to question 4.4.
* No > Please skip question 4.2 and continue to question 4.3.

**UU Guidance**

Consider the potential risks in handling your data, whether these are confidential or privacy-sensitive data or not. Think about loss of any kind of data, integrity of your data, usage by persons not in your project group, or risks that occur when anyone mistakenly accesses and uses non-quality checked or non-processed data.

**4.2. How will you handle confidential or privacy-sensitive data?**

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.

**UU Guidance on ethics and privacy**

Consider the following principles of the General Data Protection Regulation (GDPR) to make decisions on procedural measures:

* **Transparency, lawfulness, and Fairness:** State whether you have gained consent of participants in your research which includes preserving data for verification and reuse. Also evaluate if the research is fair towards the participants, regardless of consent. Check the guide [Informed consent for data sharing](https://www.uu.nl/en/research/research-data-management/guides/informed-consent-for-data-sharing)on the RDM website.
* **Purpose limitation:** Statewhat (legal) agreements or procedures you will set up with others to arrange the access conditions to your data. Describe how you will ensure that these are not in conflict with informed consent, the purpose of the data collection, or security measures needed. You can make use of the guide [Legal instruments and agreements](https://www.uu.nl/en/research/research-data-management/guides/legal-instruments-and-agreements) on the RDM website to find out more.
* **Data minimisation:** Explain the necessity of any (personal) information that you will collect.
* **Accuracy:** How will you ensure that your data is accurate and of high quality. Data subjects have to be able to rely on the correctness of their information.
* **Storage limitation:** Explain how long information needs to be retained and at what point you will either destroy, anonymise, or pseudonymise the data.
* **Security and access limitation:** Data security is needed to prevent unauthorised access, disclosure, changes or destruction of your data. You, and other principle investigators are responsible for ensuring data security and will decide for what purpose data is collected and how data may be accessed and processed. The level of security required depends upon the nature of the data you collect or use. Confidential or privacy-sensitive data needs higher levels of security. Explain how you will prevent unauthorised access and use during storage or while transferring data (e.g. during field work). State the procedures or storage strategy you will apply, e.g. separate storage of identifiable, non-identifiable data and the key, encryption of data or computers, logging of access, physical security such as automatic locking of computers, passwords, not allowing temporary storage on any other than the secure location, etc.

Make sure that you:

* Explain whether you did a Data Privacy Impact Assessment (DPIA) or [data classification](https://intranet.uu.nl/en/data-classification) to find out appropriate measures;
* Mention if you contacted an ethical committee or RDM Support;
* Identify any formal standards that you will comply with, such as the UU policy on [Information Security Management](https://intranet.uu.nl/en/information-security-policy) or any other institutional data security policies which are in place.

For more information you can use the guides [Legal instruments and agreements](https://www.uu.nl/en/research/research-data-management/guides/legal-instruments-and-agreements),  [Policies, codes of conduct, laws](https://www.uu.nl/en/research/research-data-management/guides/policies-codes-of-conduct-and-laws) and [Informed consent for data sharing](https://www.uu.nl/en/research/research-data-management/guides/informed-consent-for-data-sharing) on the RDM website

**Example answer**:

1. I have asked consent of all participants for the purpose of the investigation, the handling and storing of the data, and the possibility of sharing of the data after the project has ended under strict rules (see informed consent file X).

We will not collect more personal data than necessary, which will be the following information (list Y). Where possible, we will anonymise any identifying information at the earliest possible point (list Z). We will not store the data with the remaining identifiable information unencrypted unless in secure environments. The key to encrypted information where necessary is known to the project members and is stored elsewhere from the encrypted data. We do not store the data on portable devices.

The data is as a default stored in Yoda, an infrastructure developed at Utrecht University. This infrastructure provides an integrated collaboration and (long-term) storage environment. The internal IT security department of Utrecht University has cleared Yoda for storing data classified as sensitive, using the BIR data-classification.

2. I have explained and dealt with the ethics issues in my ethics paragraph for the ERC grant I obtained, this can be found in file X.

**4.3. What measures will be taken to comply with security requirements and mitigate risks?**

Consider necessarymeasures for the availability, integrity and confidentiality (authorisation of access) of your data.

**UU Guidance**

Make sure that you:

* Consider the availability of your data. How important is uninterrupted access? Does your storage location comply to this?
* State measures to avoid others from tampering with your data. Note down any computer system security and physical data security measures you will use. State measures to restrict access to unauthorised persons.

**Computer system security**

The computer you use to consult, process and store your data, must be secured in the following ways:

* Use a firewall;
* Install anti-virus software;
* Install updates for the operating system and software;
* Only use secured wireless networks;
* Use passwords and do not share them with anyone. Do not use passwords on your UU computer only, but also on your laptop or home computer;
* If necessary, secure individual files with a password.

**Physical data security**

With a number of simple measures, you can ensure the physical security of your research data:

* Lock your computer when leaving it for just a moment (Windows key + L);
* Lock your door if you are not in your room;
* Keep an eye on your laptop;
* Transport your USB stick or external hard disk in such a way that you cannot lose it;
* Keep non-digital material which should not be seen by others, in a locked cupboard or drawer.

**Data classification**

If you are unsure if you have taken enough and appropriate security measures, you are advised to do [the data classification of Utrecht University](https://intranet.uu.nl/en/data-classification).

**Example answer**

We will use standard computer system security and physical data security measures, such as automated locking of computers, firewalls, anti-virus software which are supplied by Utrecht University.

**4.4. To whom will access to your data be granted/restricted***?*

Please describe who will have access when to what parts of your data, including who has authority to grant access.

**UU Guidance**

A number of individuals may require access to your data, possibly with different privileges to read, write, update or delete. You might need remote access to your data yourself, if you are not at the university or working from more than one location.

The scope, actions permitted, conditions and procedures under which your data may be made available to other researchers during and after your research are determined by you as the Principal Investigator. Specify who gets access to what part of the research data, and at what time, under what conditions, and what actions are permitted. Others may include your research director, a specific colleague from the project, the data manager, your supervisor, your funder, a student assistant, a collaborator, a research partner organisation, a reviewer, or people who are just interested.

Also describe who grants access on behalf of the Principal Investigator, if this is delegated.

**Example answer**

For non-sensitive data: During the project, all project members J.B, V. van den H., K. de L. have access to all data within the project during the project. If they wish to grant (temporary) access to others, they should seek approval from all project members by motivated request.

For privacy-sensitive data: Student assistants will collect the data, and they will be asked to sign a confidentiality agreement (see form X). Access to the data will after that be restricted to members of the research team. The key to the identifiable data will be accessed only by X. van P., who is the Principal investigator. Persons requesting the data should, after granting access, sign a data transfer agreement (see agreement form Y), in which is described for what purpose the data can be used, how to store, and how long they can have the data. The principal investigator can decide whether requirements are met. This task flows to the data manager of the faculty [name] if the PI is no longer able to provide the approval.

**PRESERVE AND SHARE**

1. Data selection and preservation
2. Data availability for reuse

# 5. Data selection and preservation

Research data underlying your publication needs to be preserved for (at least) 10 years so it can be consulted for integrity, verification, reproducibility and transparency purposes. The policy of Utrecht University and many research funders is to have as much of the research data available for reuse. Take this into account and elaborate on the necessary actions for reuse in part 6: Data availability for reuse.

**Please answer the following questions**:

5.1. Which data should be preserved and/or shared?

5.2. How and where will you keep your data for the long term?

**More information**

If you want more information on data selection and preservation you can check the [RDM website of Utrecht University](https://www.uu.nl/en/research/research-data-management/guides/storing-and-preserving-data).

**5.1. Which data should be preserved and/or shared?**

Describe the content of the data package you will preserve for the long term and indicate how it will be possible to reproduce your findings.

**UU Guidance**

According to [University policy framework for research data](https://www.uu.nl/sites/default/files/university_policy_framework_for_research_data_utrecht_university_-_january_2016.pdf), your data needs to be stored for at least 10 years. This is in line with The Netherlands Code of Conduct for Scientific Practice. Decide what data to preserve and what data to dispose of after the end of your project.

Whether data is worth to preserve for the long term depends on the requirements or policies of your discipline, the type of data created, if it is possible to preserve it, and whether further work or publications will be based on the data. For data that by their nature cannot be re-measured, efforts should be made to retain them indefinitely.

Data that needs to be preserved should be stored in a ‘data package’ with all relevant describing information needed to reproduce findings (e.g. documentation files such as code, methods).

A retention period of at least 10 years may not always be viable or realistic. Sometimes data needs to be destroyed. This may have different reasons (contractual, legal or regulatory).

Also, data could already be preserved by others and is available elsewhere;

It can be cheaper or easier to reproduce the data than it is to preserve it;

There is too much raw data and processed data alone are sufficient for reproducibility.

Make sure that you:

* Describe which criteria you will use to decide which data has to be preserved for long-term access;
* Indicate which data and supportive files (like the necessary software to open the files) you will keep for at least 10 years;
* State if any data must be destroyed and if so, state which data and why.

**Example answer**

I’ll make sure that with the appropriate documentation the content can be understood by others to reproduce findings in the associated publications. The data package will contain: the methods and materials to collect the data, the raw data, the script to come to processed data, the processed data, the scripts leading to tables and figures in the publication, a codebook with explanations on the variable names, a ‘read me’ text with an overview of files included and their content and use.

In those situations where the raw data is too privacy-sensitive or large, I will only keep the processed data. This applies to the processing of audio recordings to transcripts. A random quality check of some of the conversions from raw to processed data by the data manager will ensure that the data was processed correctly.

**5.2. How and where will you keep your data available after your research?**

Explain where you will preserve your data, and how procedures are applied to ensure the survival of the data for the long term.

**UU Guidance**

The long term is a minimum of 10 years. A practical solution is to preserve your data package in a certified long-term data repository where it will also be findable and easily shareable with others (see also 6. Data availability for reuse). A good data repository or long-term data archive will replicate your data to more locations and check for decay of files or the file carrier.

If you will not use an established data archive or repository, your data management plan should demonstrate which resources and systems will be in place to enable the data to be curated effectively beyond the lifetime of the project.

Consider how your data will be preserved beyond the lifetime of your project and make sure that you:

* Indicate how long data and supportive files should be preserved;
* Specify at what location your data will be preserved;
* State how the data will be accessible during the retention period, and who is responsible for providing access;
* Consider file formats that allow for uninterrupted accessibility during the retention period;
* Indicate any additional costs to prepare data for preservation and how you will cover these;
* If applicable, indicate how you will meet the charges of data repositories.

**Example answer**

1) The data will be kept for at least ten years. They will be stored in Yoda for i-lab. Yoda is an infrastructure developed at the Utrecht University and provides for an integrated collaboration, secure and (long-term) storage environment. The data will be preserved in a vault where the data are kept safe and cannot be tampered with.

2) The data will be kept for at least ten years. They will be stored in the faculty data server. In this data server, the data is replicated and checked for bit rot regularly. The data is accessible by issuing a request to the data manager at [name]@uu.nl. In principle, the data can be obtained by anyone, only for the purpose of verifying the publication that was based on this. A data use agreement has been set up to formalize this. The use of the faculty data server for archiving is free of charge for researchers of the faculty.

# 6. Data availability for reuse

In many cases, your research data is also suitable for future reuse, by yourself or others.

By publishing your research data Findable, Accessible, Interoperable, Reusable, you make it available to the scholarly community, who can study and build upon it. Your work will become more complete, and your research data itself visible, gaining a chance to be noticed and cited, adding to your impact.

**Please answer the following questions**:

6.1. What secondary use of your data is intended or foreseeable?

6.2. Where will you make your data available?

6.3. What access and usage conditions will apply?

**More information**

If you want more information on data availability for reuse you can check the [RDM website of Utrecht University](https://www.uu.nl/en/research/research-data-management/guides/publishing-and-sharing-data).

**6.1. What secondary use is intended or foreseeable?**

Explain what reuse of your research data you intend or foresee, and what audience will be interested in your data. This will help you decide what data you will make available for reuse, where you will make the data available and how.

**UU Guidance**

At the end of your research project, you are required by the Utrecht University Policy on Research data to share your research data with others within as well as outside Utrecht University, as far as reasonably possible. Many funders also require this. This is because research data can be a valuable resource for reuse.

Make sure that you:

* Identify what part of your data is interesting for reuse by others;
* State the purpose(s) the data can be reused for, and what audience will use it;
* Mention any reasons why your data is not suitable for reuse.

**Example answer**

Our research data will be of general interest to researchers working in methane emission quantification. For others to be able to interpret the data, the context in which the data was produced is provided in the form of extensive materials and methods. The raw data, processed data, and results will be useful for others as a benchmark, to compare effect sizes and design new experiments. The raw data can be processed in meta-analyses. Or, it can be used as training material for education.

**6.2. Where will you make your data available?**

Explain where you will make your data findable and available to others.

**UU Guidance**

By publishing your data, you make it available to the scholarly community, who can study and build upon your work. Your work will become more visible. Where you make your data available strongly influences the chance your data are discovered and reused. Your choice also determines the permanence of the availability of your datasets and whether it is securely stored.

Make sure that you:

* Check if there is a discipline-specific data repository that fits your research data. An international list of data repositories is available via re3data. Some publishers provide lists of recommendations e.g., PLOS ONE recommended repositories; there is also the Data Deposit Recommendation Service for humanities researchers. Verify that your data gets a permanent link (like a Digital Object Identifier (DOI) and check if the repository has some quality standard for preserving your data such as the Data Seal of Approval or an ISO standard. When looking for a data repository that fits your data best you can also get advice from RDM Support. For quick advice on repositories, you can check IT solutions on the RDM website.Consider using the services provided by Utrecht University for making your data available for reuse like DataverseNL or Yoda;
* State how you plan to advertise your published data to your audience (by newsletter, community website, press release, attending seminars or conferences, etc.);
* Reflect on the possibility to publish your data in a peer-reviewed Data Journal. These journals offer descriptions of scientifically valuable datasets and facilitate possible reuse.
* If you do not propose to use an established data archive or repository, you should demonstrate that resources and systems will be in place to enable the data to be curated and shared effectively beyond the lifetime of the project.

**Example answer**

1. I will publish my data in Yoda, where it will obtain a DOI and be findable via DataCite, where the metadata is registered.

2. I will publish my data in the ArrayExpress repository (www.ebi.ac.uk/arrayexpress/). The repository does not have a data seal of approval, nor issues persistent identifiers. However, it is very established, international, and a standard go to as a source of information in the field of functional genomics. ArrayExpress is indexed by Thomson Reuters Data Citation Index and by SCOPUS. Datasets submitted directly to ArrayExpress are known as good sources for reuse as they are curated by a team of specialist biological curators. Data is collected according to MIAME standards.

**6.3. What conditions for access and usage will apply?**

State when your data will be available for reuse. Also specify if any restrictions or special conditions apply in accessing or using your data. State the kind of license or usage agreement you will use for your data.

**UU Guidance**

Once people find your data, it is important it’s clear to them under what conditions they can either access or reuse the data. If this is not clear, people may not be able to legally use your data.Define when you will share your data. Research funders expect a timely release. They typically allow embargoes, but they do not allow prolonged exclusive use. Consider strategies to minimise restrictions on sharing. These may include anonymising or aggregating data, gaining participant consent for data sharing, gaining copyright permissions, and having a limited embargo period using data formats that can be used by many others.

Make sure that you:

* Consider the possible licenses or terms of use and specify them. Note that when you deposit in a data repository, repositories offer one or more options to choose from;
* Name the conditions to access your data: open access, restricted access, embargo (if so, until when?); If you need to restrict access to certain communities or apply data sharing agreements, you need to make this explicit.
* State how you will organise administrative charge for access, if applicable.

Consider the following licensing and access options:

* Public domain, meaning you can use CC0 (Creative Commons Zero);
* An open license (using a Creative Commons license) such as CC BY:
* An embargo on the data until …[date];
* Restricted access with terms negotiated with requestors as needed. Please state who has authority to grant access (also after you leave the university);

**Example answer**

1. In the chosen repository, the data will be available under a CC0 license, with terms of use that require the user to cite the data. To minimise barriers for reuse, the data is available without restrictions to access, immediately after the publication of the last paper.
2. As the data is privacy-sensitive, we publish the descriptive metadata in the data repository, with a description of how a data request can be made (by sending an email to the faculty data manager at [name]@uu.nl). The data is stored as described in 5.2. If granted, a data usage agreement is signed by the receiving party (annex A.)