

Data management section NWO – UT guidance

(Last updated: 19-01-2023)

Guidance by UT:

For [NWO proposals](#), researchers need to fill in a data management section. This template of the data management section that needs to be handed in for NWO proposals contains additional guidance by the UT and various example answers that can help to fill in the data management section.

Currently, NWO has various versions for this data management section. Please check whether exactly the same questions are needed for your specific NWO proposal.

The example answers are given for inspiration, but they should not just be copied to your own data management section. The exact answers to the following questions might differ for each research project. Please contact the **data steward of your faculty** for additional help and feedback.

1. Will this project involve re-using existing research data?

- Yes: From my own or a collaborator's prior research.
- Yes: Publicly available data.
- No: Have you considered re-using existing data but discarded the possibility? Why?

If no, please briefly explain why; if yes, state any constraints on re-use of existing data if there are any.

NWO encourages the re-use of existing data wherever possible. Explain which existing data you will re-use and state any constraints on re-use of existing data if there are any.

Guidance by UT:

To discover data relevant for your research, please have a look at the [following portals and/or data providers](#) ('Searching scientific information', under heading 'Data Sources' open the table).

The advantages of reusing findable and high-quality data are:

- possibility to test new or alternative hypotheses and methods of analysis (enables the exploration of topics not envisioned by the initial investigators and encourages diversity of analysis and opinion);
- combination of datasets from multiple sources through meta-analysis promotes new types of research and permits the creation of new datasets, information, and knowledge;
- facilitation of collaborations within and outside your discipline and/or institution, possibly leading to co-publications

Example 1:

'Yes: From my own or collaborator's prior research.' and *'Yes: Publicly available data'* chosen.

We will use data that has been previously collected in our lab. No constraints regarding re-use of this data exist. In addition, we will re-use at least two datasets from 4TU.ResearchData ('DOI dataset A', 'DOI dataset B'). These datasets both have a CCO license. This means that there are no constraints regarding the re-use of these datasets.

Example 2:

'No: Have you considered re-using existing data but discarded the possibility? Why?' chosen.

Previous research (data) helped to develop our research idea. However, a thorough search of repositories did not result in any previously collected datasets that can be directly used to answer our research questions. Therefore, we will collect various datasets ourselves.

2. Will data be collected or generated that are suitable for reuse?

- Yes: Please answer questions 3 and 4.
- No: Please explain why the research will not result in reusable data or in data that cannot be stored or data that for other reasons are not relevant for reuse.

Guidance by UT:

The preferred answer to this question is 'Yes'. The basic principle is that data should be open to the public whenever possible and closed when necessary. Answering 'Yes' does not mean that all data should be openly available for reuse. You can choose 'Yes' even if only part of your data is suitable for reuse. For example, in some cases, raw data cannot be made open directly due to e.g. privacy issues, but after data anonymization/pseudonymization, the processed data can be made openly available.

In some cases you have to answer 'No'. Closure, temporary closure (embargo) or limited/restricted access can be justified by, for instance, arrangements with research funders, the General Data Protection Regulation (GDPR, in Dutch: Algemene verordening gegevensbescherming, AVG), public safety, and/or intellectual property rights of third parties.

Example 1:

'Yes: Please answer questions 3 and 4.' chosen.

→ No additional explanations are needed.

Example 2: commercial confidentiality

'No: Please explain why the research will not result in reusable data or in data that cannot be stored or data that for other reasons are not relevant for reuse.' chosen.

Due to collaborations with a commercial partner all measurement data is saved in a specific format that can only be opened with tools developed at this company. We are not allowed to convert this to another format, and/or to make any of the data available for re-use due to agreements with the commercial partner.

Example 3: privacy protection

'No: Please explain why the research will not result in reusable data or in data that cannot be stored or data that for other reasons are not relevant for reuse.' chosen.

In this project, political opinions will be collected by having interviews with people in key positions in the government in different European countries. Since the data is confidential and cannot be anonymized without making it less useful, we have made agreements with the participants that the data will only be used in this specific research project and will not be shared or reused.

3. After the project has been completed, how will the data be stored for the long-term and made available for the use by third parties? Are there possible restrictions to data sharing or embargo reasons? Please state these here.

Guidance by UT:

NWO and UT policy is that data should be 'open if possible and closed if necessary'. If the data cannot be made open, at least you should make the data traceable and available in a persistent and permanent way.

In the light of Open Science and scientific integrity, sustainably archiving your static data and providing access is crucial. Data repositories and institutional data archives let you digitally archive your (meta)data to keep your research verifiable, replicable and reusable for the long term. We recommend using a trusted repository, we point out two: [4TU.ResearchData](#) for technical and natural sciences data, and [DANS Easy](#) for humanities, health sciences, social and behavioural sciences, oral history and spatial sciences. These trusted repositories received the [Core Trust Seal](#), which guarantees reliable, citable data that can be found, accessed (clear rights and licenses) and used in the long term, even if the hardware and software become obsolete. Both repositories allow for different access options (public, embargo or restricted access) and different licenses can be chosen (e.g. [CC licenses](#)). For UT researchers, depositing data up to 1 TB per year in 4TU.ResearchData and 50 GB in DANS Easy in total is free of charge. Next to the two data repositories, you can also choose to archive research data and other supporting materials related to a research project in the UT data archive Areda. Please contact your faculty's data steward for support in this case.

There are also other (international) repositories that cater to your field of research, check the [List of repositories with a Core Trust Seal](#) or check [Re3data](#) (be aware that it has a proper quality standard, like a Core Trust Seal). You could use this license selector to choose an appropriate license: <http://ufal.github.io/public-license-selector/>. When reusing software and/or data from others: please take into account the licenses that were applied to the reused software/data.

Example 1:

Datasets that do not contain personal data will be uploaded to DANS EASY which is a trusted repository and provides a DOI for each dataset. Data will be stored there for at least 10 years. Self-written code for analyses will be publicly available under an MIT license. Other datasets will be publicly available under a CC-BY license. One dataset will be made publicly available (CC-BY license) after an embargo period of 2 years due to commercial interests of our project partners.

Datasets containing personal data will be archived in the internal archiving solutions at the University of Twente for at least 10 years in compliance with the GDPR regulations and are only available to the principal investigator and the researcher who collected the data. These datasets will never be made available to others, however, the metadata will be made available in the Research Information System (Pure) of the University of Twente.

Example 2:

As explained under question 2, (meta)data cannot be made available (for re-use) due to commercial interests. It is ensured that all data will be stored at the partner company for at least 10 years after the end of the project.

4. Will any costs (financial and time) related to data management and sharing/preservation be incurred?

- Yes: Then please be sure to specify the associated expenses in the budget table of this proposal.
- No: All the necessary resources (financial and time) to store and prepare data for sharing/preservation are or will be available at no extra cost.

Guiding by UT:

Data management and sharing activities can be costly in terms of resources and time, therefore costs for data management made during a research project can be included in a proposal's budget. Think of costs related to:

- data acquisition, processing or analysis
- data storage
- anonymization or transcription of data
- curation of data before sustainable archiving
- people, equipment, infrastructure and tools to manage, document, organise, store and provide access to the data.

You can consult this [Guide Research Data Management and Costs](#) to estimate costs associated with data management. UT-facilities/services without costs for employees are: UT network drive (storage space in Project and Organization directory depends on the number of employees in the group/department), [BMSLab storage](#) (for BMS associated projects; storage space depends on project), Surfdrive (max. 500 GB), OneDrive (max. 1 TB), Google workspace (max. 100 TB). Check the [decision tool](#) for more information about the different storage solutions and costs.

Example 1:

'No: All the necessary resources (financial and time) to store and prepare data for sharing/preservation are or will be available at no extra cost.' chosen.

For the lab devices used in this project, many steps are already automatized and relevant documentation will be added automatically to the raw data (text files with experimental settings etc.). Further documentation of the experiments and data will be performed by the PhD who will work on this project. This will not take a large amount of time as notes are already taken in the lab notebook during the experiments, and the total amount of data (incl. the number of files) is rather low. Therefore, no additional costs are budgeted for data documentation.

We will only use a rather low amount of storage (~100GB) which is available within our research group and paid by the central funds of our research group. Also, sufficient storage is available in the fridge in our lab for physical samples. In addition, every researcher from the University of Twente can upload up to 1TB of data each year for free to 4TU.ResearchData.

Example 2:

'Yes: Then please be sure to specify the associated expenses in the budget table of this proposal.' chosen.

This project consists of seven research organizations and will produce a large amount of data (>10 TB). Additional costs are involved for storing and archiving this data at our project partner's. In addition, documenting all the data and code for this project, and integrating all the data files from multiple project partners for final archiving is time intensive. Therefore, additional costs are included for a technician who is helping with the documentation and data integration. The exact costs are specified in the budget table of this proposal.