

<i>date</i>	12 March 2019
<i>creator</i>	TU Delft Library
<i>contact</i>	Faculty Data Stewards
<i>e-mail</i>	<a href="mailto:datastewards@tudelft.nl">datastewards@tudelft.nl</a>

## **Guidelines for a data paragraph in a NWO project proposal**

These guidelines are aimed at helping researchers writing a data paragraph in an NWO research proposal. If you want us to go over your draft paragraph, don't hesitate to [contact your faculty Data Steward](#) and feel free to ask for help!

For proposals, [NWO](#) would like you to answer four questions on research data management.

If funding is awarded, you'll be asked to submit a longer data management plan in which you will have to describe in more details how you will manage your data. For now, you can be brief; the main message to get across is that you're aware of the importance of good data management, have already put some thought into how to manage data during the project, and are willing to make your data available for reuse by other researchers after completion of your project. If there are valid reasons not to share research data publicly, these should be explained in the data paragraph.

We will go through the four questions. Each question is accompanied by an example originating from a real data paragraph.

### **1. Will data be collected or generated that are suitable for reuse?**

NWO states<sup>1</sup>: NWO understands 'data' to be both collected, unprocessed data as well as analysed, generated data. This can be in all conceivable formats; digital and non-digital (for example samples, completed questionnaires, sound recordings, etc.). In principle NWO assumes that there is a widely held view within different disciplines about which data are relevant to store for reuse. The importance and the value of reuse on the one hand and the costs and feasibility of data storage on the other should be in reasonable balance with each other and have a bearing on the volume of the data to be stored. Not every research project yields reusable data.

**Yes:** We recommend to provide a brief description of the data you are going to collect or generate. If you will be using any datasets from third party sources (e.g. datasets from the Central Bureau for Statistics), this also needs to be explained. In addition, consider mentioning any models, code or software which will be created, generated or used to process the data. Subsequently, explain if the datasets will be suitable for sharing and re-use. Note that explaining that your datasets will be useful for re-use by the community, will make your proposal look more impactful and, possibly, more likely to be funded. Moreover, by making your data available for other researchers, you demonstrate adherence to [the Netherlands Code of Conduct for Research Integrity](#).

**No:** Then explain why the research will not result in reusable data.

There can be valid reasons to answer 'no', mostly related to personal research data (which cannot be anonymised), commercially sensitive data, or other forms of confidential data. Datasets that contain

personal data, which cannot be anonymized without making the data less useful are typically not suitable for public sharing. However, such datasets can still be shared by depositing in repositories providing restricted access to data (such as [DANS EASY](#)). Another constrain to data sharing might come from working with data from commercial partners, where researchers need permission to make the data available for reuse.

‘No’ can also apply to data that can be easily replicated, and where it would take more time and effort to document and store it, than to replicate it.

#### Example

The original experimental data and programming source code are extremely important for the future research to check, validate and reuse the results.

During the research, the data and expected volume will be in four aspects:

- Original experimental data, 200 GB
- Programming source code and archive, 300 GB
- Numerical simulation results, 300 GB
- Documentation and articles: 200 GB

The research data gathered in this project will be described in a Data Management Plan in the first four months of the project under the guidance of the Faculty Data Steward and updated during the course of the project.

## **2. Where will the data be stored during the research?**

NWO states<sup>1</sup>: During the research NWO prefers digital data to be stored in a safe place where it is accessible to others with the permission of the researcher.

Here you demonstrate that you have thought about safety and security of your data: how are you going to prevent data loss (think computer crashes, backups etc.) and how will you ensure that your data is inaccessible to people who shouldn't be able to access your data (especially important if you work with commercially sensitive or personal data).

Depending on your needs there are several options. Visit TU Delft [TOPdesk](#) for all available storage facilities (the most common recommendation is to use High Availability data storage for research) or [contact your Data Steward for help](#).

*Please note that if your project will generate any types of confidential data (e.g. personal or commercially sensitive research data), this section will need to be appropriately adapted. Please [contact your Data Steward for help](#).*

#### Example

During the course of the research project, all data will be stored on TU Delft network drives (Project Storage space) maintained and automatically backed up by TU Delft ICT. This storage will be used for sharing data with authorised collaborators inside and outside of TU Delft. The Faculty Data Steward (in collaboration with the Faculty's ICT manager) will provide additional advice, as needed, on data storage during the research project.

## **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? To whom will the data be accessible?**

NWO states<sup>1</sup>: After the research the data should preferably be archived at a national or international data repository. If that is not possible, the data should be archived by the institutional repository. Confidential, privacy-sensitive or competition-sensitive data might require special forms of storage or

limited access. Solely storing the data on computers or external media (e.g. USB flash drive, CD, DVD of hard disks) is in general too risky and can therefore, in principle, not be approved by NWO.

In the Netherlands, we have two national repositories: 4TU.Centre for Research Data, a certified repository for technical-scientific research data located at the TU Delft Library, and DANS EASY, that mostly contains data from the social sciences and humanities. 4TU.Centre for Research Data offers 1TB per year of free storage to all TU Delft researchers and also offers them up to [5,000 EURO towards the costs of preparing data for deposit](#). Alternatively, you could also choose an international repository that caters to your field of research. Other data repositories can be found via [Repository Finder](#) or get in touch with your [Data Steward for help](#).

Choosing a data repository at this stage may seem a bit premature, but data repositories can have specific requirements: they only accept certain file formats, they only allow certain licences on data sets, etc. - those are things you don't want to discover at the end of your project, since fulfilling those requirements then could cost you a lot of time that you wouldn't have had to spend if you had known them in advance.

*Please note that if your project will generate any types of confidential data (e.g. personal or commercially sensitive research data), this section will need to be appropriately adapted. Please [contact your Data Steward for help](#).*

#### Example

All datasets will be published at [4TU.Centre for Research Data](#), which is a trusted and certified research data repository (Data Seal of Approval certification). All datasets will be accompanied by metadata, compliant with DataCite metadata schema, to ensure that all datasets are findable and accessible online. All datasets will be publicly available to anyone for re-use under an open licence. Every dataset will be also assigned a Digital Object Identifier (DOI), to make them citable and persistently available. 4TU.Centre for Research Data preserves the datasets for at least 15 years, maintaining their integrity and authenticity, in accordance with the [4TU.Centre for Research Data preservation policy](#).

#### **4. Which facilities (ICT, (secure) archive, refrigerators or legal expertise) do you expect will be needed for the storage of data during the research and after the research? Are these available?**

The answer to this question of course highly depends on what kind of research you are planning and what kind of data you are going to generate or collect. If your research data includes samples and physical collections that need to be archived for verification or reproducibility purposes, check with your department if they have physical storage spaces for samples used during research.

*Please note that if your project will generate any types of confidential data (e.g. personal, or commercially sensitive research data), this section will need to be appropriately adapted. Please [contact your Data Steward for help](#). Also please contact your Data Steward in case more storage and archival space is needed than described below.*

#### Example

The project will not lead to collection/creation of any confidential research data and therefore will not require use of any additional facilities on top of those discussed above. All the facilities discussed above are readily available to use. We do not expect the total data to exceed 5TB and therefore there will not be any costs for data storage on local servers. Up to 1TB of data can be archived free of charge (per year, per TU Delft author) on 4TU.Centre for Research Data. We do not expect to exceed this.

## Costs

*Note, this part is not included in the Data Paragraph question, but should be incorporated into your data management costing part of the grant. If you have any questions about preparing this part of your grant, please [get in touch with your Data Steward](#).*

You might use the [Research Data Management Costing Tool](#) to help you prepare this section (note, the tool is in a beta version – any issues noticed should be reported to [datastewards@tudelft.nl](mailto:datastewards@tudelft.nl)).

In general, researchers should seek to recover the direct costs of managing research data from the research funder.

- The costs associated with research data management should be considered at the earliest opportunity.
- Costs of data management might include:
  - Personnel cost:
    - Cost of hiring a dedicated Data Manager for the project to assist with data management tasks in the project:
      - If new Data Managers are hired, they should be budgeted at scale 9 (if relevant research background is not necessary) or at scale 10 (if it is necessary that Data Managers have a prior research experience).
    - Time associated with the preparation of research data for archiving and publication including the relevant documentation.
  - Hardware and infrastructure costs (if infrastructure other than TU Delft's is used):
    - Costs of access to any specialist infrastructure, such as High Performance Computing.
    - Cloud computing costs.
  - Software costs (if TU Delft does not have the licences):
    - Purchase of licences for software to support good data management, such as Electronic Lab Notebooks, or project management software.
  - Publishing costs:
    - Data publication, if there is a charge associated with depositing in a data archive.
    - Publication of papers about datasets or software in dedicated journals.

### More info and links

- Guidance on [Data paragraphs](#)
- Guidance on [Data Management Plans](#)

<sup>1</sup> <http://www.nwo.nl/en/policies/open+science/data+management>