

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

**These guidelines are aimed at helping researchers writing a data paragraph in a NWO research proposal. If you want us to go over your draft proposal, don't hesitate to contact us and feel free to ask any follow-up questions.**

For proposals, NWO would like you to answer four questions on research data management. The referees and the assessment committee can give feedback on your answers, but this section of your application is *not* included in the decision about the awarding of the funding.

If funding is awarded, you'll be asked to submit a longer data management plan in which you describe in more detail 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.

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?**

*Yes: We recommend to provide a brief description of the data you are going to collect or generate. Then answer questions 2 to 4.*

*No: Then 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.*

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.). NWO only requests storage of data that are relevant for reuse. 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.

Data that can be reused by other researchers don't have to be generated or collected again. Creating or generating reusable data can also benefit you as a researcher: reuse of your data by colleagues, either in your own field of research or in other disciplines, means increased citations. Moreover, by making your data available for other researchers, you demonstrate adherence to *the Netherlands Code of Conduct for Academic Practice*.

However, there can be valid reasons to answer 'no', mostly to do with ethics and property rights. Datasets that contain personal data are harder to share with others, since anonymization of personal data isn't always possible without making the data less usable for research, and if you use data from commercial partners you'll need their permission to make the data available for reuse.

'No' can also apply to data that is easily repeatable and where it would take more time and effort to document and store than to repeat again.

### Example

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

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 Research Data Management Plan (RDM plan) in the first four months of the project and updated during the course of the project.

Most of publications and articles will be published in international well recognized journals, for example, Journal of Cement and Concrete Research, Journal of Cement and Concrete Composites and Journal of Materials and Structures. Or published in open access journals, for example, Journal of Reviews on Advanced Materials Science, Journal of Advances in Materials Science and Engineering and Journal of Civil Engineering and Urbanism.

## **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 sensitive/personal data).

Depending on your needs there are several options. Use the '[Open Science Guide](#)' to get an overview of all the available options.

### Example

Datasets will be stored both in DataverseNL and TU Delft ICT server space (SURFdrive and project data storage at TU Delft) during the project. Back-up of the data will be done daily and at two locations in NL.

Advice on proper storage and archiving as well as data description, sharing and publishing will be given by the TU Delft Library during the course of the research project.

### Further example

During research we will store the data at our own faculty server. These data are backed up on a daily basis. Advice on secure storage is provided by the ICT Department.

Once the project has been completed, the data will be transferred to the data archive of

4TU.Centre for Research Data (short, 4TU.ResearchData) to make it available to a wider audience.

The research data gathered in this project will be described in a Research Data Management Plan (DMP), which will be drafted upon acceptance and updated during the course of the 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 or 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.ResearchData, a repository for technical-scientific research data located at the TU Delft Library, and DANS EASY, that mostly contains data from Alfa and Gamma sciences. But you could also choose an international repository that caters to your field of research. Other data repositories can be found at [re3data](#).

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 licenses 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.

#### Example

The research data gathered in the project will be securely stored at 4TU.ResearchData (<http://researchdata.4tu.nl>), which is a Trusted Digital Repository for technical-scientific research data in the Netherlands. Each dataset deposited will be provided a Digital Object Identifier, or DOI, that facilitates discoverability, accessibility, and re-use. 4TU.ResearchData has been included in the

DataCite search engine and Thomson Reuters Data Citation Index (DCI) to aid data discovery. Data collections, structured databases, processed data and data representations will be stored for 15 years after the end of the project. Raw data underlying these data sets will be kept available but not accessible online.

Research data that is not privacy sensitive will be open access available through the data repository mentioned above as far as this is compatible with and does not infringe IP requirements of the companies with which I collaborate in this project. Details on possible restrictions to data access will be discussed with the company partners and described in further detail in the data management plan.

**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 needs to be archived for verification or reproducibility purposes, check with your department if they have physical storage spaces for samples used during student research. Consult the 'Open Science Guide' as referred to in question 2 or seek advice from your Faculty IT Manager or Library staff.

They also should be able to help you 'translate' the facilities you need into costs to put in your proposed budget. The project budget must include the total cost of archival storage including specific equipment and facilities needed for the proposed storage time.

The costs for data management are eligible for funding and can therefore be entered in the project budget. We recommend you to make these costs explicit.

Example

Legal expertise is needed for the storage of the data. They are available in the TU Delft Library. Computer used as server during research and data storage after finishing the project are taken into account in the budget plan.

ICT facilities for data storage are considered to be issues such as data storage capacity, bandwidth for data transport and calculating power for data processing.

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