date	12 March 2019
creator	TU Delft Library
contact	Faculty Data stewards
e-mail	datastewards@tudelft.nl

Guidelines for a data paragraph in a Horizon 2020 project proposal

These guidelines are aimed at helping researchers writing a data paragraph in a Horizon 2020 research proposal. If you want us to go over your draft proposal, don't hesitate to contact your faculty Data Steward and feel free to ask for help!

In Horizon 2020 the Commission has launched a **flexible pilot for open access to research data (ORD pilot)**. The pilot **aims to improve and maximize access to and re-use of research data generated by Horizon 2020 projects.**

In previous Work Programmes, the ORD Pilot was limited to some areas of Horizon 2020.

As of 2017, the Open Research Data pilot is extended to cover all thematic areas of Horizon 2020 by default. However, the Commission recognizes that some research data cannot be made open and applies the principle of *'as open as possible, as closed as necessary'*. The possibilities to opt out remain, i.e. for privacy-related reasons, to respect intellectual property rights, or whenever the public release of the data might jeopardize the project's main objective. In this context, the guidelines for data management for Horizon 2020 have been updated.

This means that all project proposals must include a section on research data management which is evaluated under the criterion 'Impact'. Applicants must provide a short, general outline of their policy for data management in which they answer the following questions:

- What types of data will the project generate/collect?
- What standards will be used?
- How will this data be exploited and/or shared/made accessible for verification and re-use? If data cannot be made available, explain why.
- How will this data be curated and preserved?

Applicants should also ensure resource and budgetary planning for data management and include an initial Data Management Plan (DMP) as a deliverable at month 6 (at the latest) into the proposal.

We will go through the four questions:

What types of data will the project generate/collect?
What kind of data will the project collect or generate, and to whom might they be useful later on. This includes, firstly, the data and metadata needed to validate results in scientific publications and, secondly, other curated and/or raw data and metadata that may be required for validation or reusability purposes.
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. 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 available for reuse.

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.

What standards will be used?
In this section you provide answers on questions such as "What is the data about?", "Who created it and why?", "In what forms is it available?", "What software is being used?".
Metadata enables data to be found and understood, ideally according to the particular standards of your scientific discipline. Provide reference to existing suitable standards in your discipline. If these do not exist, provide an outline on how and what metadata will be created.
FAIRsharing (https://fairsharing.org) or the DCC website (http://www.dccc.ac.uk/resources/metadata-standards) can be consulted to identify suitable disciplinary metadata standards or seek advice from your faculty Data Steward.

How will this data be exploited and/or shared/made accessible for verification and re-use? If data cannot be made available, explain why.
By default as much of the resulting data as possible should be made publicly available. Note that explaining that your datasets will be suitable for sharing and 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*.

Provide a description of how data will be shared, including: access procedures; embargo periods (if any); explanation of dissemination strategy and necessary software and other tools for enabling re-use; and definition of whether access will be widely open or restricted to specific groups. In case the dataset cannot be shared, the reasons for this should be mentioned accordingly (e.g. protection of personal data, intellectual property, commercial, privacy-related, security-related).

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.

How will this data be curated and preserved?
You should select a data repository that will preserve your data, metadata, and offer support guidelines for data standards and licensing. Give an indication of how long the data should be preserved, what its approximated end volume is, what the associated costs are, and how these are planned to be covered.

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.

Once a project has had its funding approved and has started, you **must submit a first version of your DMP** (as a deliverable) within the first 6 months of the project. It is advisable to contact your faculty Data Steward when writing the first version of a Data Management Plan.
Projects that opt-out of the ORD pilot are still encouraged to submit a DMP on a voluntary basis.

More info and links:

- Official EC Guideline:
  - http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (incl. DMP template)

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

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).

- 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.

- Guidance on Data paragraphs and data management plans