5 Things to check when reviewing Data Management Plans (DMPs) - *Guidance for supervisors*

PhD candidates must submit a Data Management Plan (DMP) as part of their Go/No-Go meeting, in accordance with the TU Delft Research Data Framework Policy. Below are some considerations for evaluating DMPs.

Please be aware that you bear the final responsibility for the data and code of the PhD candidate you're supervising; It is in your interest to check that they are following good data management practices.

Tip: If you have any questions or would like to seek further advice, contact your faculty data steward(s).

---

1. Was a data steward consulted?

It is advised, although not required, to reach out to the faculty data steward(s) for support on a DMP. This guide gives general advice on best practices, but it is not exhaustive. The data stewards are experts in research data management who can provide you with more comprehensive advice.

If your faculty data steward(s) hasn't yet been consulted on the DMP, ask the PhD candidate to reach out to them by using the "Request Feedback" button on DMPOnline, or by contacting them directly through the contact information on the contact page.

Tip: TU Delft's Data Management Plan template is more comprehensive and better integrated into our internal services.

2. Is the PhD Candidate using good data storage practices?

TU Delft offers two main solutions for the storage of data and code: **Project Drive** (for research data) and **GitLab** (for code management). Both solutions are backed up and maintained by TU Delft, so you don't have to worry about data and code being accidentally lost. For a comparison between **GitLab** (TU Delft version) or other open-source repository managers based on Git, check here.

If another device must be used for data storage, it is advised to use a TU Delft laptop/workstation. As with any solution, it is important to ensure that:

- The necessary security measures, such as file access permissions and (where appropriate) encryption, are enforced so that access is only given to project members. This is to minimise the risk of unauthorised access and data leaks.
- There is a proper backup strategy in place to prevent data loss.

Some researchers also use **SURFdrive or OneDrive (TU Delft version), but they are not recommended for long-term storage** of research data. SURFdrive and OneDrive are personal cloud storage spaces - which means that once the person leaves, not only them but also you, as their supervisor, will lose access to their data.

Tip: Check if you can easily find and get access to the data and code managed by the PhD candidate
=> Is all the data and/or code where the DMP said it would be?
=> Is there data and/or code in the storage location that isn't in the DMP?

3. Is the data properly documented?

Do you think that the data and code are **properly documented**? For example, can you easily figure out what each file is? If you open some of the files, are they properly labelled and described? Crucially, **what if the PhD candidate leaves TU Delft tomorrow**? Would you be able to make sense of the data and how the research was done?It is recommended that all data sets and software come with a text (or markdown) file that explains how the accompanying files are structured, how they can be read, and provides some context on their use.

You can find some templates and guidance for documentation here:.

- Software / Code README (TUD DCC)
- Data README (4TU.ResearchData)
- Data / Code / ML models READMEs (TUD AE)
- Recommended coding practices (TUD AE)

Tip: Browse through the data/code of the PhD candidate and see if they make sense to you.

4. Is the PhD candidate working with confidential data?

In short: is the PhD candidate aware of the considerations, requirements and best practices regarding confidential data?

Confidential data can contain:

- **personal data** which can allow the identification of living individuals
- **commercially-confidential information** (such as data or code that discloses something directly about a patentable invention, or data belonging to a third party)
- Information related to **national security**, export control regulations and dual use

If the PhD candidate is working with any type of confidential data, the data needs to be securely stored and only authorised people should have access to the data. In addition, if the project involves human subjects and/or working with personal research data, it is required to submit an application to the Human Research Ethics Committee.

In principle, if the PhD candidate is working with any type of confidential data, they will most likely need to use data storage and code management recommended by TU Delft (Project Drive, GitLab).

If the PhD candidate is working with confidential data from third parties, check any contractual agreements to see if the data and code is managed accordingly.

In addition, for some types of confidential data, especially if there is a risk of accidental data release (data temporarily stored on laptops or USB drives), data might have to be **encrypted**. In some cases, confidential data (such as criminal record databases, or documents relating to national security) supplied by third parties may not be allowed to leave their servers, or they may require that the data only be accessed on-premises. In these cases, contact your faculty data steward(s) for advice.

Tip: When accessing the data, check who else has access rights to that data. Browse through the data: what would happen if the data was accidentally released? Does it require any extra security measures, such as encryption?

5. Is there a clear strategy for publishing and archiving research data?

Research data is the evidence underlying research findings, and therefore should be published and archived for long term preservation where possible. TU Delft Research Data Framework Policy requires **all PhD candidates who started on or after 1 January 2019 to deposit research data (and code) supporting their theses in a research data repository before they can graduate** (unless there is a valid reason why this is not possible).

Possibilities that exist:

- data/code can be openly published
  (everyone can see it, access it and reuse it; data/code gets a persistent identifier);
- data/code is published with restricted access due to confidentiality
  (everyone can see the metadata, but people have to request access to access it and reuse it under specific terms; data/code still gets a persistent identifier);
- data/code is kept in-house at TU Delft available to TU Delft research team and collaborators only
  (only applicable for data/code that cannot be shared, no persistent identifier).

Questions to address:

- Did you already have a discussion about this with the PhD candidate?
- Did you agree what, when, and how the data (and code) will be published and archived?
- What data is *useful* to publish? What does someone else need to reproduce the results in your paper(s)?
- Will there be any difficulties with publishing or archiving the research data (and code)?
- What licence will be the most suitable for the data (and code)?

In principle, for most research done at TU Delft, it is suitable to make the data and code underpinning research findings available in a data repository. Typically, this is done no later than when publishing the related papers, theses or reports. TU Delft has a dedicated data repository, 4TU.ResearchData, where all TU Delft researchers can deposit up to 1TB of data (and code) per year (per researcher) free of charge. Discipline specific repositories might also be suitable (the faculty data steward(s) can advise).

If you or the PhD candidate anticipate any problems with data sharing you can ask your faculty data steward(s) for advice.

Tip: When deciding what data (and code) to share, think about the following questions:
=> What data (raw, processed, final data) (and code) are necessary to validate and reproduce/replicate research findings?
=> What data (and code) can't be re-generated? (such as weather observations)?
=> Are there any data publishing or archiving requirements from your funder?
=> Consider what you wished other authors would have added to their publications.