RDM Checklist Doctoral Defence - Form B

Form B is a declaration that all the final checks and compliance with requirements have been fulfilled. For doctoral candidates who started their Doctoral Programme from 1 January 2019 onwards, this also includes the requirement of uploading their research data and code supporting their dissertation to a data repository following the FAIR principles.

This checklist intends to be a guide and/or a reminder of:

- the final things related to the publication of research data and software you should discuss with the PhD candidate before you sign the Form B.

- the final things to check on the published research data and/or software/code before you sign Form B.

If you, as a supervisor have not been discussing the publication of research data and/or software/code during the PhD trajectory with the PhD candidate, we recommend you to check one of these resources to get familiar with the requirements of the TU Delft Research Data Framework Policy and what is necessary in preparation for that:

- These pages can guide the PhD candidate on Publishing their Research Data.

- On the RDM website you find general information about Publishing Research Data and Publishing Research Software.

- On this guide you find a summary of an information session 'Publishing requirements for data and code for TU Delft PhD candidates' run by the Faculty Data Stewards together with the Graduate School.

To get started:

- Request to the PhD candidate a list of the research data and/or code that has been already published, ideally they will provide the DOIs.

    - Ideally, data and code underlying scientific articles should have been published at the time of the publication of the corresponding papers.

- You can use the Data Management Plan (DMP) as an inventory to check what was planned to be published and what was not intended to be published.

    - Ideally the list of published datasets/code provided by the PhD candidate should align with the agreements in the DMP.

- You can recommend PhD candidates to use the DMPonline's 'Research outputs' feature to create a list of datasets and code, if they have access to the corresponding DOIs.

For each published dataset/software/code:

- Check that they are assigned a persistent identifier (usually a DOI) in the data repository where they are published.
- Check if the data files are well structured (data, methods, and outputs should be clearly separated; the raw data should be separated from the processed data; the computational environment should be specified).
- Check if all the files can be opened (and are not corrupted).
- Check if the data is stored in open or standard format to ensure long-term preservation. Check the preferred file format at the 4TU.ResearchData or Library of Congress Recommended Formats for detailed info.
- Check if there is a README file, as a minimum documentation, and that can be opened in an open file format, e.g. .txt or .md.
- If applicable, check if any additional documentation is available along with the data and/or code (e.g. a data dictionary, Electronic Lab Notebook, Readthedocs, GitHub Wiki, etc.).
- Check that you are listed on the documentation as a contact person.

Check that the README file contains a minimum of relevant information about the data and/or software/code, such as:

- Title
- Author(s) names, affiliations, contact info and ORCID
- List of all the files and accompanying documentation
- Methodology of data collection and analysis, including (if applicable) any setting or protocol, assumptions and software (including version number) used to analyse the data
- Description of the data and/or code, the file naming and structure of data
- References to related articles, with DOI
- Licences and/or copyright information
- Recommended way of citing

- Any other information that helps readers to better understand the data and/or code
- Some good templates of README files are found here:
  - Software / Code README (TU Delft DCC)
  - Data README (4TU.ResearchData)
  - Data / Code / ML models READMEs (TU Delft AE)

If during the project, personal data or confidential data was published with restricted access or as a metadata-only record:

- Check if the following information is included in the metadata or other accompanying documentation:
  - The reason for data being confidential
  - The data ownership
  - The contact info for requesting access to the datasets
  - The protocol for accessing the data
  - Check that you are listed on the documentation as a contact person

Check for **research data and/or software/code that can't be published**:

- Remember that valid reasons for not publishing research data and software/code could include:
  - working with confidential data and/or software
  - working/collecting personal data that cannot be anonymized or pseudonymized
  - working/collecting data or developing software that can be severely misused or falls under special regulations, for example, export control
- Make sure to discuss with the PhD candidate **what will happen with the research data and software/code that can't be published**. In general there are two options, but if in doubt you should consult with the Faculty Data Steward:
  1. Data/Software/Code can be internally **archived**: Project Data (U:) Drive/Staff - Umbrella is the recommended solution. Make sure that the ownership or administrative right of the Project Data (U:) Drive has been transferred to you before you sign Form B or at the latest before the PhD

candidate leaves TU Delft. Also remember to transfer the data or ownership to any shared cloud storage (e.g. a shared folder on SURFDrive).

2. Data/Software/Code can be **deleted**. If data/code is irrelevant or too sensitive/confidential to be safely and legally preserved. Make sure you have clear agreements on that with the PhD candidate. Those agreements can still be recorded in the DMP and kept as documentation.

- If the PhD candidate has developed software in a private GitLab or GitHub repository, make sure you have access to it before the candidate leaves TU Delft.

Final reflections:

- Do not miss the opportunity to talk with the PhD candidates about their learnings regarding research data and software management along the project. You can use their experience to better support the next candidates that you will be mentoring:
    - Ask what was most challenging and what were the major learnings about RDM practices?
    - What would they have done differently?
    - What could you have supported them more with?
    - What RDM learnings will they take with them into their next job?