

Data management plan for UL doctoral students

Danijel Skočaj

University of Ljubljana
Faculty of Computer and Information Science

Ljubljana, 7. 4. 2022



Univerza v Ljubljani



- Purpose: to consider how the data collected or generated in the course of the PhD research will be preserved and made accessible to others
- Considering FAIR principles
- Regulated by the Rules and Regulations for Doctoral Studies at the University of Ljubljana
 - Article 50
 - Applied for the students enrolled from the 2021/22 academic year onwards
- Guidelines and templates for research data management planning for PhD students at the University of Ljubljana
 - About the data
 - Storage and protection of the data during research for the doctoral thesis
 - Long-term data availability and storage
- Two goals:
 - To be aware of the importance of the appropriate handling of data and to plan data management accordingly
 - Not to impose unnecessary additional paper work to doctoral candidates

Findable: Data has rich metadata and unique identifier

Accessible: Data can be easily downloaded

Interoperable: Metadata use an accessible and standard language

Reusable: Data is well described

Research data management

Article 50

RULES AND REGULATIONS FOR DOCTORAL STUDIES AT THE UNIVERSITY OF LJUBLJANA

Research data generated and collected for the needs of a doctoral dissertation must be published or otherwise accessible in such a way that allows their visibility, access, interoperability and the possibility of renewed evaluation and use. The doctoral candidate shall submit research data to a data repository, data centre or research data archive, which shall satisfy the principle of verifiability, transparency and open science. As a priority the research data shall be sent to the sectoral national or international data centres intended for specific types of data, or to the UL Repository.

The doctoral dissertation shall state where the data are accessible and how they can be accessed. Exceptions in the sharing of data shall be justified where they involve personal or sensitive data, or where there are reasons for protecting intellectual property or for non-disclosure of vulnerable areas, groups or species. In the case of implementing justified exceptions to data sharing, the doctoral candidate shall ensure an appropriate method of protecting the data and limiting access to such data in agreement with the data centre. In this case at least freely available metadata must be generated for the catalogue of the data centre, so as to indicate clearly where and under what conditions the research data are accessible.

1. When submitting an application for approval of the doctoral dissertation proposal
 - DMP draft

Article 36

A doctoral candidate shall submit an application for approval of the doctoral dissertation proposal to the responsible body in accordance with the rules and regulations of the UL member.

• • •

- a draft of the research data management plan in accordance with Article 50 of these Rules;

• • •

2. When presenting the results of the research work at the end of the doctoral study
 - Updated DMP (almost final)

Presentation of the results of research work before completion of the doctoral dissertation

Article 43

When the supervisor, potential co-supervisor and doctoral candidate are of the opinion that research work is in its final stage, and when it is possible to submit a draft of the conclusions in accordance with the set objectives or hypotheses or research questions, the doctoral candidate shall submit a draft of the doctoral dissertation to the members of the DSC, and shall present the results of their research work, with an emphasis on the main findings and contribution to science, and also the updated research data management plan in accordance with Article 50 of these Rules.

• • •

3. When submitting the final version of the doctoral dissertation
 - Final DMP

Article 45

A doctoral dissertation shall include:

- • •
- the final draft of the research data management plan, and other possible annexes;
- • •

Data management plan (DMP)

DMP Draft - when submitting your doctoral dissertation proposal

| |
|--|
| <p>Name of the doctoral student:</p> <p>Doctoral programme and scientific field:</p> <p>Proposed title of the doctoral dissertation:</p> |
| <p>Type of data and methods used for data collection or production</p> <ol style="list-style-type: none">1. What data will be collected or produced?2. How will new data be collected or produced and/or how will existing data be re-used for the purposes of your doctoral thesis?3. Will you be dealing with sensitive data? If yes, how will you ensure compliance with ethical requirements when producing and/or creating data? |
| <p>How data will be stored and protected during research for a doctoral thesis</p> <ol style="list-style-type: none">1. How will data be stored?2. If you will be dealing with sensitive data, how will you keep it safe and secure? (Move to the next question if not applicable) |
| <p>Long-term data availability and storage</p> <ol style="list-style-type: none">1. In which data repository will you store the data for the long term after the completion of the research work and make it accessible in accordance with the requirement of Article 50 of the Regulations on Doctoral Studies of the UL?2. Do you plan to restrict access to the data for a certain period? If yes, please explain the reasons for this (e.g. for intellectual property or patent protection, or other reasons). |

Data management plan (DMP)

DMP - at the presentation of the results of the research or at the submission of the doctoral dissertation

| |
|---|
| <p>Name of the doctoral student:</p> <p>Doctoral programme and scientific field:</p> <p>Title of the doctoral dissertation:</p> |
| <p>Type of data and methods used if new data was collected or produced</p> <ol style="list-style-type: none">1. What data have you collected and/or produced?2. How did you collect and/or produce new data and how did you use existing data for your doctoral thesis?3. Have you been dealing with sensitive data? If yes, how have you ensured compliance with ethical measurements when collecting and/or producing data? |
| <p>How the data was stored and protected during research for a doctoral thesis</p> <ol style="list-style-type: none">1. How did you store the data?2. If you have worked with sensitive data, how have you ensured that it was kept safe and secure? (Move to the next question if not applicable) |
| <p>Long-term data availability and storage</p> <ol style="list-style-type: none">1. In which data repository will you store the data for the long term after the completion of the research work and make it accessible in accordance with the requirement of Article 50 of the Regulations on Doctoral Studies of the UL?2. Do you plan to restrict access to the data for a certain period? If yes, please explain the reasons (e.g. for intellectual property or patent protection, or other reasons). |

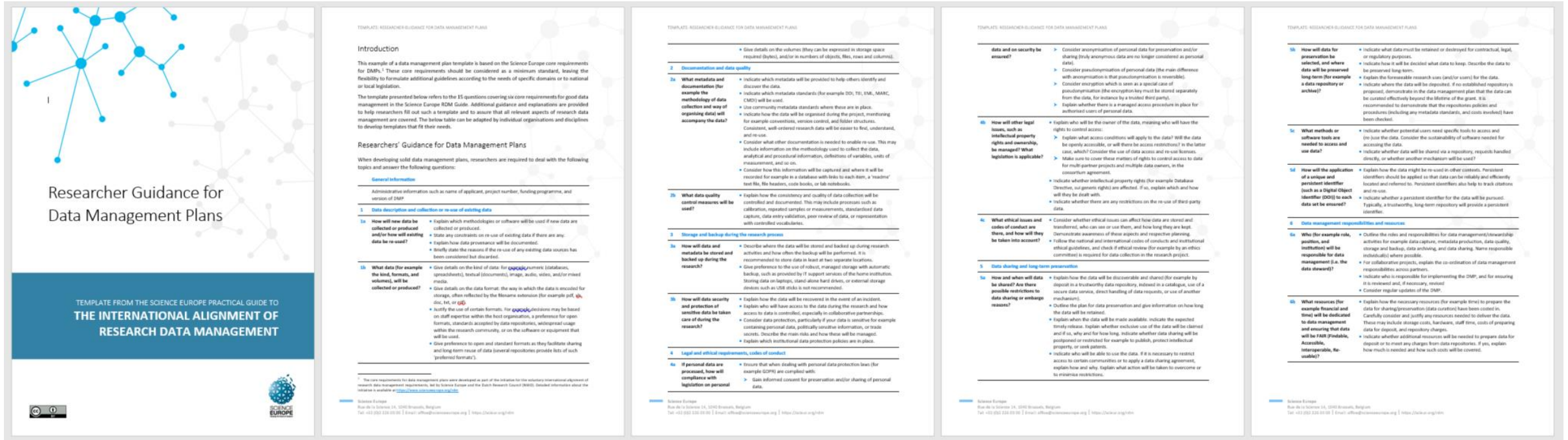
- https://www.uni-lj.si/study/doctoral_school/research_data_management/
- https://www.uni-lj.si/doktorska_sola/raziskovalni_podatki/

- **Type of data and methods used for data collection or production**
 - What data will be collected or produced?
 - How will new data be collected or produced and/or how will existing data be re-used for the purposes of your doctoral thesis?
 - Will you be dealing with sensitive data? If yes, how will you ensure compliance with ethical requirements when producing and/or creating data?
- **How data will be stored and protected during research for a doctoral thesis**
 - How will data be stored?
 - If you will be dealing with sensitive data, how will you keep it safe and secure? (Move to the next question if not applicable)
- **Long-term data availability and storage**
 - In which data repository will you store the data for the long term after the completion of the research work and make it accessible in accordance with the requirement of Article 50 of the Regulations on Doctoral Studies of the UL?
 - Do you plan to restrict access to the data for a certain period? If yes, please explain the reasons for this (e.g. for intellectual property or patent protection, or other reasons).

- **Type of data and methods used if new data was collected or produced**
 - What data have you collected and/or produced?
 - How did you collect and/or produce new data and how did you use existing data for your doctoral thesis?
 - Have you been dealing with sensitive data? If yes, how have you ensured compliance with ethical measurements when collecting and/or producing data?
- **How the data was stored and protected during research for a doctoral thesis**
 - How did you store the data?
 - If you have worked with sensitive data, how have you ensured that it was kept safe and secure? (Move to the next question if not applicable)
- **Long-term data availability and storage**
 - In which data repository will you store the data for the long term after the completion of the research work and make it accessible?
 - Do you plan to restrict access to the data for a certain period? If yes, please explain the reasons?

Further guidance for DMP

- More details:
<https://scienceeurope.org/our-priorities/research-data/research-data-management/>



Researcher Guidance for Data Management Plans

TEMPLATE FROM THE SCIENCE EUROPE PRACTICAL GUIDE TO THE INTERNATIONAL ALIGNMENT OF RESEARCH DATA MANAGEMENT

1. **Data description and collection or re-use of existing data**

1a. How will new data be collected or produced and/or how will existing data be re-used?

- Explain which methodologies or software will be used if new data are collected or produced.
- State any constraints on re-use of existing data if there are any.
- Explain how data provenance will be documented.
- Briefly state the reasons if the re-use of any existing data sources has been considered but discarded.

1b. What data (for example the kind, formats, and volumes), will be collected or produced?

- Give details on the kind of data (for example, spreadsheets, databases, spreadsheets, textual documents), image, audio, video, and/or mixed media.
- Give details on the data format, the way in which the data is encoded for storage, often reflected by the filename extension (for example pdf, doc, txt, or xls).
- Justify the use of certain formats, for example decisions may be based on staff expertise within the host organization, a preference for open formats, standards accepted by data repositories, widespread usage within the research community, or on the software or equipment that will be used.
- Give preference to open and standard formats as they facilitate sharing and long-term reuse of data (several repositories provide lists of such "preferred formats").

2. **Storage and backup during the research process**

2a. How will data and metadata be stored and backed up during the research?

- Describe where the data will be stored and backed up during research activities and how often the backup will be performed. It is recommended to store data in at least two separate locations.
- Give preference to the use of robust, managed storage with automatic backup, such as provided by IT support services of the home institution, storing data on laptops, hard drive hard drives, or external storage devices such as USB sticks is not recommended.

2b. How will data security and protection of sensitive data be taken care of during the research?

- Explain how the data will be recovered in the event of an incident.
- Explain who will have access to the data during the research and how access to data is controlled, especially in collaborative partnerships.
- Consider data protection, particularly if your data is sensitive for example containing personal data, politically sensitive information, or trade secrets. Describe the main risks and how these will be managed.
- Explain which institutional data protection policies are in place.

3. **Legal and ethical requirements, codes of conduct**

3a. If personal data are processed, how will compliance with legislation on personal data?

- Ensure that when dealing with personal data protection laws (for example GDPR) are complied with.
- State informed consent for preservation and/or sharing of personal data.

4. **Data sharing and long-term preservation**

4a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

- Explain how the data will be discoverable and shared (for example by deposit in a trustworthy data repository, indexed in a catalogue, use of a secure data service, direct handling of data requests, or use of another mechanism).
- Outline the plan for data preservation and give information on how long the data will be retained.
- Explain when the data will be made available, indicate the expected timely release. Explain whether exclusive use of the data will be claimed and if so, who and for how long. Indicate whether data sharing will be postponed or restricted for example to publish, protect intellectual property, or seek patents.
- Indicate who will be able to use the data. If it is necessary to restrict access to certain communities or to apply a data sharing agreement, explain how and why. Explain what action will be taken to overcome or to relieve restrictions.

5. **Data management responsibilities and resources**

5a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

- Outline the roles and responsibilities for data management/tutorship activities for example data capture, metadata production, data quality, storage and backup, data archiving, and data sharing. Name responsible individuals where possible.
- For collaborative projects, explain the co-ordination of data management responsibilities across partners.
- Indicate who is responsible for implementing the DMP, and for ensuring it is reviewed and, if necessary, revised.
- Consider regular updates of the DMP.

5b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Reusable)?

- Explain how the necessary resources (for example time) to prepare the data for sharing/preservation (data curation) have been collected in. Carefully consider and justify any resources needed to deliver the data. These may include storage costs, hardware, staff time, costs of preparing data for deposit, and repository charges.
- Indicate whether additional resources will be needed to prepare data for deposit or to meet any changes from data repositories. If yes, explain how much is needed and how such costs will be covered.

- Start thinking about the data management at the beginning of the doctoral study
- Consult your mentor
- The template might be used (or you can use also some other form of presenting the required information)
- Adapt the document and content to the specifics of your research discipline
- Answers to the questions are expected to be brief and concise
- Manage the data according to the FAIR principles

