Checklist 'Prepare your dataset'

SUMMARY

DATA SELECTION

- ☐ Consider whether Data@UNIMI is the right platform for uploading your research data. It can host data that are at the basis of publications (or PhDs theses) or research projects, and, generally, data that have been processed and/or represent the final stages of a research process: **make sure that you select relevant data and keep uncomplete data elsewhere. Please, check the guidelines:** [Data@UNIMI or Drive?](#)

- ☐ Ensure the data has not been published yet or received a DOI (or other persistent identifier) in other (trusted) repositories.

- ☐ You can store **30 GB per year**. If your dataset exceeds this limit, contact [Data@UNIMI support.](#)

LEGAL AND INTELLECTUAL PROPERTY ISSUES

Research at the University of Milan is carried out in accordance with existing contractual agreements, legislation, regulations and guidelines.

- ☐ **Check whether you have the rights to share the data**, considering:
  - ○ Existing contractual agreements
  - ○ Rights of third parties
  - ○ Confidentiality obligations
  - ○ Informed consent forms
  - ○ Ethical approvals
  - ○ Funding requirements

- ☐ **Check compliance with GDPR in case you are using** personal data:
  - ○ Contact "Ufficio di Segreteria Tecnica del Responsabile Protezione Dati" in order to register the processing of personal data
  - ○ Anonymize/pseudonymize the data. For tools on anonymization, refer to Amnesia, the OpenAIRE data anonymization tool.
  - ○ Consult with the DPO support office in case of questions.

- ☐ Notify Data@UNIMI support before submitting or disclosing potentially patentable or commercially useful data, in accordance with the Unimi Regulations on Intellectual Property Rights to Research Results.

DATA QUALITY AND DATA CLEANING

- ☐ **Check the quality of your data** using some useful tools and **provide the maximum level of FAIRness in your data to allow its reproducibility and reuse**.
- ☐ Check errors in the data.
- ☐ **Label your data coherently** (e.g. data headers, file naming, etc.): organize the name following a project standard, in such a way that the file name is somehow explaining the content.
- ☐ Preferably choose open formats or generally accepted standard formats.
- ☐ Check the size of your files: each individual file must not exceed the size of 5 GB. The zip file upload limit is 10 GB, and it can contain a maximum of 1.000 items.
- ☐ Organize your data consistently: here are some guidelines to look at on How to share data related to a publication and on How to structure a dataverse for your project.

DOCUMENTATION

- ☐ **Provide all the necessary information to understand your data,** document fully its provenance, the context, and tools of its production and/or collection.
- ☐ **Complete in the most accurate way as many fields of metadata as possible** to describe your work. Particularly, pay attention to:
  - ○ Title – mandatory
  - ○ Author(s) – mandatory
  - ○ Point of contact – mandatory
  - ○ Description (make it clear how data were created and eventually elaborated, providing all details to allow replication and reuse) – mandatory
  - ○ Subject – mandatory
  - ○ Keywords
  - ○ Related publication
  - ○ Funding information
  - ○ Time period and/or Date of collection
  - ○ Data type
  - ○ Software (eventually)
  - ○ Data source
- ☐ **Create a README file**: it is mandatory when the data format is not FAIR, but it is always recommended. Describe your collection methodology in the documentation to guarantee data quality and so that data can be replicated and/or combined in different contexts.
- ☐ Explain acronyms.
- ☐ Use well-known terminology and domain standards specific to disciplines.
- ☐ If you have many files, provide a file list (with file names, descriptions of the content and of any connections between the files).
- ☐ **Use as open as possible and as close as necessary license.**

FURTHER HELP

- ➢ Didn't find what you are looking for? Check the guide or the UNIMI data management website.
- ➢ Still have some questions? Contact Data@UNIMI support.