
Giorgio's Plan

A Data Management Plan created using DMPonline.be

Creator: Giorgio Boccarella

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

ID: 199238

Last modified: 30-04-2023

Giorgio's Plan

Application DMP

Questionnaire

Describe the datatypes (surveys, sequences, manuscripts, objects ...) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

In this research project, we will generate new data using mathematical modeling and simulations. No personal data will be involved in this project. The data types we will use include simulation output files (e.g., time series data), mathematical models (e.g., equations and code), and research manuscripts. Data formats will primarily consist of plain text files (.txt), comma-separated value files (.csv), and programming language-specific formats (e.g., .mat for MATLAB, .ipynb for Python notebooks).

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

1.
 - Responsible person: Prof. Piet van de Berg
 - Storage capacity/repository:
 - During the research: Local storage on lab PCs, with regular backups to external hard drives and/or cloud storage services.
 - After the research: Data will be deposited in an open-access platform, such as Mendeley data to ensure long-term preservation and accessibility.

What's the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

I do not wish to deviate

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

- As our research does not involve any personal or sensitive data, there are no specific ethical concerns related to the data management highlighted in the ethics questionnaire. Consequently, no particular security measures are required for the generated data

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

- Data sharing: The generated data will be made publicly available through an open-access platform, enabling other researchers to access, reuse, and build upon our findings.
- Data documentation: We will provide comprehensive metadata and documentation to facilitate the understanding and reuse of our data by others.
- Version control: We will use version control tools (e.g., Git) to track changes in code and model development, ensuring a clear history of revisions and enabling reproducibility.

Giorgio's Plan DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

- Not applicable

Giorgio's Plan GDPR

GDPR

Have you registered personal data processing activities for this project?

- Not applicable

Giorgio's Plan

FWO DMP (Flemish Standard DMP)

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

1. Name: Simulation Output

Description: Time series data generated through computational simulations of the mathematical models.

Data type: Newly generated

Content: Observational (simulation-based)

Format: .csv, .txt

Volume: ~2 GB

2. Name: Mathematical Models

Description: Equations, algorithms, and code used for developing and running the simulations.

Data type: Newly generated

Content: Experimental (mathematical modeling)

Format: .mat (MATLAB), .ipynb (Python notebooks), .m (MATLAB scripts), .py (Python scripts)

Volume: ~500 MB

3. Name: Research Manuscripts

Description: Drafts, figures, and final versions of research articles, conference papers, and presentations related to the project.

Data type: Newly generated

Content: Text and images (research outputs)

Format: .docx (Microsoft Word), .pptx (Microsoft PowerPoint), .pdf (Portable Document Format)

Volume: ~1 GB

4. Name: Literature Resources

Description: Articles, books, and other relevant resources used to inform and support the development of the mathematical models and simulations.

Data type: Reused

Content: Text (literature)

Format: .pdf (Portable Document Format)

Volume: ~2 GB

These estimates represent the upper limit of the volume of data, and actual data sizes may vary depending on the specific needs and progress of the research project.

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

Not in the current plan

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, ...)? If so, please comment per dataset or data type where appropriate.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in

the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

To ensure that the data generated and collected during the research project remains understandable and usable both in the present and future, we will adopt a comprehensive approach to data documentation and metadata creation. This approach will involve the following:

1. README Files: For each dataset, we will create a README.txt file that provides an overview of the data, its organization, and the file formats used. This file will include information on data collection, processing, and any modifications made, as well as descriptions of variables, units, and any necessary context for understanding the data.
2. Codebooks: For datasets containing variables or codes, we will create a Codebook.tsv file that details variable names, descriptions, units, and codes, along with explanations of any abbreviations or acronyms used. This codebook will ensure that users can easily understand the meaning and context of each variable in the dataset.
3. Electronic Lab Notebooks (ELNs): Throughout the research process, we will maintain ELNs to document the development of mathematical models, simulation parameters, and data analysis procedures. This documentation will provide a chronological record of the research process, including details on any assumptions, decisions, and modifications made during the project.
4. Commented Code: We will ensure that all code used for mathematical models, simulations, and data analysis is well-commented and organized. This will facilitate the understanding of the code by others, allowing them to reproduce the research or adapt the code for their own purposes.
5. Data Management Plan (DMP): We will maintain an up-to-date DMP that outlines our data management approach, procedures, and responsibilities throughout the project. This plan will ensure that all team members are aware of their data management roles and will provide a reference for any necessary documentation or metadata creation.
6. Metadata Standards: When depositing data in open-access platforms or repositories, we will follow established metadata standards (such as Dublin Core, DataCite, or discipline-specific standards) to ensure that our data is easily discoverable and interoperable with other datasets.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- No

3. Data storage & back-up during the research project

Where will the data be stored?

1. Local Storage: Data will be stored locally on lab PCs. This allows for quick access and efficient processing during the research project. Regular backups will be performed to ensure data safety.
2. KU Leuven Data Cloud: To provide additional security and accessibility, data will be stored on the KU Leuven Data Cloud. This cloud storage service offers a secure, centralized location for data storage, enabling easy access and collaboration among team members.
3. Mendeley Data: As an open-access platform, Mendeley Data will be used to store and share the research data publicly. This platform ensures long-term preservation and accessibility of the data, allowing other researchers to access, reuse, and build upon our findings.

How will the data be backed up?

1. Local Backups: Data stored on lab PCs will be regularly backed up to external hard drives.
2. KU Leuven Data Cloud: The data stored on the KU Leuven Data Cloud is automatically backed up and versioned by the service provider. This offers an additional layer of protection for the data, ensuring that even if the local copies are lost or damaged, a secure backup is available in the cloud.
3. Offsite Backups: In addition to the local and cloud backups, we will periodically create offsite backups of critical data by storing copies on external hard drives or other secure storage media. These backups will be stored at a separate location to protect against risks such as natural disasters, theft, or damage to the lab facilities.
4. Version Control: For code and other version-sensitive materials, we will use version control systems (e.g., Git) to track changes and maintain a history of revisions. This approach allows for easy recovery of previous versions of the code or data, in case of accidental deletion or corruption.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- Yes

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

To ensure that the research data is securely stored and protected from unauthorized access or modification, we will implement the following security measures:

1. Access Control: Data stored on lab PCs, network-attached storage devices, and cloud storage services will be protected by strong, unique passwords. Access to these systems will be granted only to authorized personnel working on the project. User accounts will have the minimum necessary permissions to perform their tasks, following the principle of least privilege.
2. Encryption: Sensitive data, including backups stored on external hard drives or offsite storage media, will be encrypted using strong encryption algorithms. This ensures that even if the physical storage media is lost or stolen, the data remains secure and inaccessible to unauthorized individuals.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

No extra costs.

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

Yes

Where will these data be archived (stored and curated for the long-term)?

For long-term archiving and preservation of the research data, we will use a reputable and stable open-access data repository. The chosen repository will ensure that the data remains accessible, discoverable, and reusable over time, even beyond the duration of the project. Some options for archiving the research data include:

1. Zenodo: A general-purpose open-access repository developed by CERN and OpenAIRE. Zenodo accepts a wide variety of data types and formats and provides a DOI for each dataset, ensuring long-term accessibility and citability.
2. Figshare: A cloud-based open-access data repository that allows researchers to store, share, and discover research data. Figshare supports a wide range of file formats and assigns DOIs to uploaded datasets, ensuring long-term preservation and discoverability.
3. Mendeley Data: A secure cloud-based repository by Elsevier that enables researchers to store, share, and manage their research data. Mendeley Data assigns a DOI to each dataset and supports versioning, making it suitable for long-term archiving.
4. KU Leuven Data Repository: A university-specific data repository that ensures long-term preservation and access to research data generated by KU Leuven researchers. The repository assigns a DOI to each dataset and adheres to international metadata standards, ensuring data discoverability and interoperability.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

None

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in an Open Access repository

If access is restricted, please specify who will be able to access the data and under what conditions.

Not restricted

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- No

Where will the data be made available? If already known, please provide a repository per dataset or data type.

Not clear

When will the data be made available?

During publication

Which data usage licenses are you going to provide? If none, please explain why.

Not clear at the moment

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- No

What are the expected costs for data sharing? How will these costs be covered?

None

6. Responsibilities

Who will manage data documentation and metadata during the research project?

Giorgio Boccarella and Prof. Piet van de Berg

Who will manage data storage and backup during the research project?

Giorgio Boccarella and Prof. Piet van de Berg

Who will manage data preservation and sharing?

Giorgio Boccarella and Prof. Piet van de Berg

Who will update and implement this DMP?

Giorgio Boccarella and Prof. Piet van de Berg

*