optimal and efficient sampling for rectangular approximation methods

A Data Management Plan created using DMPonline.be

Creator: Astrid Herremans

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: Astrid Herremans

Data Manager: Astrid Herremans

Project Administrator: Astrid Herremans

Grant number / URL: 11P2T24N

ID: 206835

Start date: 01-11-2023

End date: 31-10-2027

Last modified: 26-04-2024

optimal and efficient sampling for rectangular approximation methods
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)

The project is expected to generate:
- scientific software
- manuscripts: papers, conference proceedings and research notes
- presentations: research seminars and conference proceedings
- experiments: data is expected to have a mathematical or numerical nature
No personal information about individuals, groups of people or organisations will be collected or processed in the project.

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)

All data in the project is treated as stipulated in the Data Management Plan of the NUMA research section, in which the work is carried out. The plan formulates both general and specific rules and practices, specific to the context of researchers at NUMA. The responsible person in NUMA is prof. Giovanni Samaey.
The plan explains its purpose and technical aspects at NUMA related to backup, version control and archival of data. It describes proper management of code, manuscripts, experiments, literature and third party software, respectively, in detail. Individual researchers write and sign a statement in which they detail the choices for their research project.

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)

There will be no deviation of the principle of preservation of data in the project.

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)

NA

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

The data management in this project is standard in the NUMA research section.

optimal and efficient sampling for rectangular approximation methods
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.

Dataset Name	Description	New or reused	Digital or Physical	Only for digital data Digital Data Type	Only for digital data Digital Data format	Only for digital data Digital data volume (MB/GB/TB)	Only for physical data Physical volume
Own code	The code I will be writing throughout my PhD (written in Matlab / Julia)	*Generate new data*	Digital	*Simulation data*	*.m or .jl*	< 100 MB	
Computational experiments	Documents stating the outcome of computational experiments for reproducibility, this accompanies a piece of code	Generate new data	Digital	Experimental	.pdf	< 1GB	

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:

Possibly, I will re-use third-party software, including external julia packages and built-in matlab code. Used third-party software will be listed in the computational experiments (including version number) for reproducibility.

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).

As mentioned before, pieces of code can be part of computational experiments, which contain documentation on:
- Goal of experiment (a few sentences)
- Which code was used? Which version?
- The script for the experiment itself.
- The processed data (tables, figures)
- Interpretation
Also, I publish code along with published papers, containing all code needed to reproduce the experiments described in the paper. This is done via a github repository, which is referred to in the paper. This will be accompanied by a README.txt file and documentation.

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

See above.

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

Where will the data be stored?

All data will be stored on my own laptop. The meta data and computational experiments will also be stored on our departmental cloud system (NextCloud) and the code will also be stored on a version control repository (the KU Leuven gitlab server).

How will the data be backed up?

The data is backed up using the version control repository (manually) and the departmental cloud system (automatically).

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

The combined capacity of the KU Leuven gitlab for code and the departmental cloud system for documents is enough to accommodate for all the data I will generate throughout my PhD.

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

The departmental cloud system on NextCloud and the KU Leuven gitlab server are secured. I am the only one who has access. My laptop is also secured using a password.

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

The departmental cloud system (NextCloud) and the KU Leuven gitlab server are owned by KU Leuven.

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...).

The department has a system for both preserving and archiving data. Preserved data is important enough to be stored in a way that allows easy retrieval, also after the researchers that generated it have left. This data will include finished software related to manuscripts and documents related to educational tasks. Archived files can be retrieved (they are not lost), but access is not immediate. These will include all my generated data (all the code and computational experiments).

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

As mentioned before, the department has a system for both preserving and archiving data. Version controlled repositories are shared with a generic NUMA (= our research unit) user. In this way, an archived repository can be restored or preserved repositories can be accessed even in my absence to ensure continuity and accessibility of code at all times. Files on the departmental cloud-based system (NextCloud) are also shared with a generic NUMA account.

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

The cloud-based system and the version control repositories are owned by the department.

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
- Yes, in a restricted access repository (after approval, institutional access only, …)

As mentioned before, code related to published papers will be available on an open access repository on Github. Preserved code will be available on the KU Leuven gitlab, shared with the generic NUMA account, such that this can be accessed by NUMA members even in my absence to ensure continuity and accessibility of code at all times.

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

The code on the KU Leuven gitlab server will be accessible for the NUMA group.

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.

The data linked to a published paper will be made available on a designated, open access repository on Github. The other preserved data will be made available on the KU Leuven Gitlab, accessible for members in the NUMA group.

When will the data be made available?

The data linked to a published paper will be made available when the paper gets published. The preserved data on the KU Leuven Gitlab will be made available after the project is finished.

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

NA

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?

No fee is charged.

6. Responsibilities

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

Astrid Herremans

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

Astrid Herremans

Who will manage data preservation and sharing?

Giovanni Samaey

Who will update and implement this DMP?

Astrid Herremans