Exploiting combinatorial structures for algebraic and geometric decompositions

A Data Management Plan created using DMPonline.be

Creator: Fatemeh Mohammadi

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: Fatemeh Mohammadi

Grant number / URL: G0F5921N

ID: 200906

Start date: 01-08-2022

End date: 31-08-2029

Project abstract:

We will develop novel tools to solve several important real-world problems: (i) Proving safety of programs (Computer Science), (ii) Computing network reliability (Industrial Engineering), (iii) Causality (Statistics), and (iv) Geometry of particle interactions (Physics). These problems are all traditionally modeled as polynomial systems. However, given that solving a general system is very difficult, they all lack scalable algorithms. The main idea is that our applications' systems tend to have additional structural properties. Our vision is to exploit these specific properties to sidestep the difficulty of solving general systems, and obtain dedicated solution methods for realworld cases.

The most profound impact is in program verification, focusing on proving the presence/absence of bugs or vulnerabilities in code. Given the ever-increasing role of software in safety-critical operations, e.g. avionics and healthcare, it is vital to perform software verification reliably and exactly. Our results will have a huge downstream effect on every aspect of verification, ultimately leading to more secure and trustworthy software in all sorts of applications. Another focus is Network Reliability with significant applications in economics and epidemiology. We will develop novel methods to compute the reliability. Finally, we will study the Amplituhedron: a geometric object that dramatically simplifies calculations of particle interactions in Physics.

Last modified: 02-07-2023

Exploiting combinatorial structures for algebraic and geometric decompositions 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 research will not collect any data. However, it will produce software code and artifacts. Anyresulting software will be distributed using a free and open-source license and made available to the general public through long-term archives such as Zenodo and also the Ken Leuven repositories.
The results will be also published as articles in journals and proceeding of main conferences in our field.

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. Designation of responsible person (If already designated, please fill in his/her name.)
2. Storage capacity/repository
    - during the research
    - after the research

The PI (Fatemeh Mohammadi) will ensure long-term open-access publication of the results by publishing in (preferably gold)open-access venues and depositing copies of all publications on both the KU Leuven repository and arxiv.

In terms of software packages, there are a few megabyte and will be saved on KU Leuven (Computer ScienceDepartment) repositories. They will be also shared via Zenodo.

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)

Not Applicable

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)

Not Applicable

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

Not Applicable

Exploiting combinatorial structures for algebraic and geometric decompositions 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.

Not Applicable				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .csv,.pdf, .txt, .rtf, .dwg, .gml, ... • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
software package		New	Digital	software package	.pdf	<1GB	
papers/notes		New	Digital	Observation/papers	.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:

Not Applicable

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

Not Applicable

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

Not Applicable

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

Not Applicable

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 the only data that I will create in this project will be research papers and software packages. I will provide precise and easy to understand documentation with the software packages which will be available to the general public through long-term archives such as Zenodo.

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?

OneDrive for Business is a Microsoft cloud solution to securely store documents and files. Multifactor authentication with the KU Leuven authenticator app OR additional encryption will be activated to ensure the safe storage of (strictly) confidential data. Only the PI can provide access to the OneDrive for study personnel.

How will the data be backed up?

A back-up is provided via automatic version management of the files in OneDrive, maintaining up to 100 versions per file.

A second copy will be kept on the departmental level in the Computer Science Department.

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

Every member of the KU Leuven personnel benefits from a generous allocation of 2 TB data storage on OneDrive. This ample storage capacity ensures that there is more than enough room to accommodate the estimated sizes of the datasets, including papers and software packages, which do not exceed 100 GB. Moreover, the backup capacity is also fully sufficient to safeguard the integrity and availability of the stored data. We can confidently rely on the robust storage infrastructure provided by OneDrive to securely store and backup our files.

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

OneDrive ensures the utmost privacy and security for the files by employing strict access controls. Any files that we do not explicitly share remain entirely inaccessible to others. In order to provide an additional layer of protection for this particular dataset, a dedicated folder will be established, and it will be encrypted to safeguard its contents. Only the Principal Investigator (PI) and authorized collaborating researchers will be granted access to this folder, which will require the use of a unique encryption key. This ensures that only the designated individuals can view and interact with the dataset, guaranteeing its confidentiality.

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

OneDrive for Business is free for staff and students of KU Leuven.
The Department of Computer Sciences provides additional storage space for every member, where the costs are covered by the department.

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...).

Digital data: All digitally generated data will be archived for minimally 10 years after study completion, in line with the KU Leuven RDM policy.

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

Digital data: The generated research data and documentation necessary to reuse the data will be transferred to OneDrive for Business for long-term data archiving, managed by KU Leuven with automatic back-up procedures.

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

For the KU Leuven members, OneDrive for Business is free and the cost will be taken care of by the CS 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

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

Not applicable

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 KU Leuven institutional repository,
software packages which will be available to thegeneral public through long-term archives such as Zenodo. Papers will be available on ArXiv

When will the data be made available?

Upon publication of research results.

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

To foster open sharing and proper attribution, all project data that can be shared will be made available under the Creative Commons Attribution License (CC-BY 4.0). This licensing choice ensures that users who access and utilize the data are required to give appropriate credit to the original creators.

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.

- Yes

A DOI will be available through RDR, but is not yet available

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

RDR is free for KU Leuven personnel, hence, no costs are expected for data sharing.

6. Responsibilities

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

Fatemeh Mohammadi (PI)

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

Fatemeh Mohammadi (PI)

Who will manage data preservation and sharing?

Fatemeh Mohammadi (PI)

Who will update and implement this DMP?

Fatemeh Mohammadi (PI)

Exploiting combinatorial structures for algebraic and geometric decompositions GDPR

GDPR

Have you registered personal data processing activities for this project?

- Not applicable

Exploiting combinatorial structures for algebraic and geometric decompositions DPIA

DPIA

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

- Not applicable