ADVANCED MULTI-STAGE PROGRAMMING WITH COMPUTATIONAL EFFECTS

A Data Management Plan created using DMPonline.be

Creator: Tom Schrijvers

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Grant number / URL: G0A9423N

ID: 201021

Start date: 01-01-2023

End date: 31-12-2026

Project abstract:

Developer productivity and software performance are two key properties of software development that are hard to achieve simultaneously. The former requires rich layers of abstraction and highly reusable, contextindependent code. In contrast, the latter aims at code without generic overhead or indirections that is highly specialized. General-purpose optimized compilation can automatically improve the performance of high-level programs, but is still far from optimal. Much more promising are code generation techniques that can be employed in "active libraries" to offer high-level programming interfaces, and exploit knowledge of the problem domain and of the code structure to generated highly tuned code. Standard code-generation techniques (e.g., templates in C++) can be quite error-prone and hard to debug. Multi-stage programming promises to considerably lower the threshold by providing strong static guarantees with respect to well-scopedness and well-typedness at the level of the generators. This project aims to advance the state-of-the-art of multi-stage programming. In particular, we aim to develop the interaction between staging and computational effects. While most existing effort has targeted effect-free code, side-effects are essential for functionality and/or performance in most applications. We take a foundational approach, developing a formal calculus for multi-staging with support for algebraic effects and handlers, complemented by practical case studies.

Last modified: 10-07-2023

ADVANCED MULTI-STAGE PROGRAMMING WITH COMPUTATIONAL EFFECTS 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)

- Manuscripts: research papers and reports
- Source code of software prototypes built

The estimated overall storage capacity required is in the order of megabytes.

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. Tom Schrijvers
2. Copies of all the manuscripts will be stored in the KU Leuven Lirias publication database from the time of publication onwards (i.e., during and after the project). All source code will be stored in KU Leuven gitlab repositories both during and after the 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)

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)

None.

ADVANCED MULTI-STAGE PROGRAMMING WITH COMPUTATIONAL EFFECTS
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
		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 prototypes	source code of software prototypes	newly written	digital	software	source code files of the programming languages used	<100MB	
manuscripts	papers and reports	newly written	digital	manuscripts	.tex	<100MB	

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

Not applicable.

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

Not applicable.

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).

The manuscripts are written in latex and are self-documenting / require no documentation.
The course code will be documented by the manuscripts and by README files.

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?

In the KU Leuven gitlab repository for the project.
Copies of papers will also be stored in the KU Leuven Lirias publication database.

How will the data be backed up?

Lirias and gitlab manage the backups for us as part of the service.

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 gitlab repository provides at least 1GB of storage, which is plenty for this project.

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

Gitlab access is restricted to authorised parties only.

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

The storage is provided by KU Leuven at no extra cost.

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...).

All data will be preserved for at least five years.

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

The data will remain stored in the gitlab and Lirias repositories after the project.

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

The storage is provided by KU Leuven at no extra cost.

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 source code will be made available in a publicly accessible github repository and the papers in the KU Leuven Lirias database.

When will the data be made available?

Upon publication of the research.

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

The source code will be provided with an MIT license.

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?

There are no expected costs.

6. Responsibilities

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

Tom Schrijvers

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

Tom Schrijvers

Who will manage data preservation and sharing?

Tom Schrijvers

Who will update and implement this DMP?

Tom Schrijvers

GDPR

Have you registered personal data processing activities for this project?

- No

DPIA

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

- Not applicable

Created using DMPonline.be. Last modified 10 July 2023

7 of 7