DMP title

Project Name FWO DMP - Optimization algorithms for real-world 2D cutting and packing problems - DMP title

Grant Title 1S71222N

Principal Investigator / Researcher Jeroen Gardeyn

Description Cutting and packing (C&P) problems concern either packing a set of small items into a set of large objects or cutting a set of small items from these larger objects. C&P problems occur in many different fields and can range from efficiently packing pallets of vegetables for delivery to extracting the maximum value from rough diamonds. Although many C&P problems have been extensively studied since the 1970s, developing efficient solution methods remains difficult. These difficulties are compounded when problems are extended with additional constraints and/or secondary objectives, which is often the case when addressing real-world problems. Solution methods in the academic literature can, however, rarely be used for real-world problems. Indeed, existing methods are often so specific that slight changes to the problem are either impossible or greatly impairs their effectiveness. Industry thus has a strong appetite for more generic and adaptable algorithms and data structures which can cater to specific constraints and objectives, enabling them to efficiently solve their real-world problems. This project will address these needs by developing an open-source framework for 2D rectangular C&P problems, with a strong focus on generality and real-world applicability. This framework will enable industry to better model and optimize their unique problems. This will not only result in financial benefits but will also lead to a range of additional economic and environmental benefits.

Institution KU Leuven

1. General Information

Name applicant

Jeroen Gardeyn and Tony Wauters

FWO Project Number & Title

1S71222N

Affiliation

- KU Leuven

2. Data description

Will you generate/collect new data and/or make use of existing data?

- Generate new data
- Reuse existing data

Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).

1) Input data. The effectiveness of the developed algorithms will be tested on problem datasets. Both existing ones from the academic literature, newly generated, or ones from the real world. The exact format of these files will probably vary, but all of them are text based. A single file is generally <1MB, and there will be 1000s of them. Total size should not extend a couple of GB.
2) Output data from experiments. The algorithm produces solution files and logs (both text based). In addition, comparisons and analysis files will also be generated and will generally be in the form of Excel sheets (.xlsx). When archived in a compressed format (e.g. zip) the total size of all this data should not exceed a couple of GB.
3) Source code. All developed source code will be stored in git repositories. This includes code for the optimization, visualization and pre/post-processing scripts.

3. Legal and ethical issues

Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.

- No

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, add the reference to the formal approval by the relevant ethical review committee(s)

- No

Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?

- No

Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?

- No

4. Documentation and metadata

What documentation will be provided to enable reuse of the data collected/generated in this project?

1. Both result files and newly generated problem files will have accompanying documents (PDF) in which the format of the file is clearly explained.

2. For all software, README files will be added to the respective GIT repositories which will contain thorough explanations of how to build and run the software.

Will a metadata standard be used? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.

- No

All steps in the research will be documented as explained in Section 4.1 to ensure findability and reusability of the data. Furthermore, the use of git has become a de- facto standard for code-sharing and documentation.

5. Data storage and backup during the FWO project

Where will the data be stored?

1. All relevant documents on computer(s) of the researcher are automatically backupped daily to the NAS of the research group and synced to Microsoft Onedrive.

2. Code will be stored using version control system GIT on both KU Leuven's GitLab and personal repositories (e.g. GitHub).

3. Finalized results, generated instances and documentation will also be stored in GIT repositories.

How is backup of the data provided?

Software like Duplicati can make scheduled and incremental backups to many different locations (e.g. NAS).

In addition, it also enables restoring data back to the state of specific past days (depending on the configured retention policy).

KU Leuven's Onedrive will also be used in addition, which syncs continuously.

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 total required capacity will not exceed a couple GB. Both the Onedrive subscriptions (2TB), and the research group's NAS have plenty available storage capacity.

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

No significant additional costs will have to be made.

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

During the project, the data will only be stored at locations which are accessible by the relevant researchers and administrators. If required, read-only permissions can be granted to other people. Once the project is finished, the data can be made public.

6. Data preservation after the FWO project

Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).

All the data will be retained.

Where will the data be archived (= stored for the longer term)?

The data will be stored in repositories belonging to the research group on KU Leuven's GitLab servers for at least 10 years.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

Since the total size of the data will not exceed a couple of GB's, the extra cost for preservation for 5 year will be negligible.

7. Data sharing and reuse

Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

- Yes. Specify:

Some of the problem instance datasets used during this project can be provided by companies, which potentially might not agree to make their instances openly available. However, since the instances are practically anonymous, we do not expect many companies prohibiting us from making their problem instances publicly available.

Which data will be made available after the end of the project?

All previously mentioned type of data (input data, output data, source code) which can be made public will be made available.

Where/how will the data be made available for reuse?

- In an Open Access repository

When will the data be made available?

- Immediately after the end of the project

Who will be able to access the data and under what conditions?

Everyone will be able to access the data. We intend to release the source code under the GNU General Public License (GPL), which is a copyleft license.

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

No significant cost are associated with sharing the data.

We expect most of the data to be made public via platforms like GitHub (or KU Leuven's own GitLab), which are free to use.

8. Responsibilities

Who will be responsible for data documentation & metadata?

Jeroen Gardeyn (PhD student) and Tony Wauters (supervisor)

Who will be responsible for data storage & back up during the project?

Data backup will be done automatically using tools like Duplicati.

Who will be responsible for ensuring data preservation and reuse ?

Jeroen Gardeyn (PhD student) and Tony Wauters (supervisor)

Who bears the end responsibility for updating & implementing this DMP?

Jeroen Gardeyn (PhD student) and Tony Wauters (supervisor)