Data Management Plan: 1S97322N

Project name: The development of an effective control framework for large-scale autonomous robot applications.

Project identifier: 1S97322N

Principal Investigator / Researcher: Sander Teck

Description: With the advent of internet and e-commerce, companies want to store their products as close as possible to their customers in order to provide fast services. However, many high-volume e-commerce markets are located in high-wage regions. Furthermore, the short delivery times expected in e-commerce require warehouses to operate 24/7. Hence, further automation (e.g. AGVs) and optimization in distribution and manufacturing environments is necessary to provide such services. With modern technologies, robots can perform very accurate movements. However, if such a movement is not the right action at the right time, inefficiencies and wasted movements occur. To avoid this, good control policies for these robots are necessary. This research focuses on the tactical and operational level, more specifically: the vehicle scheduling, dispatching, and routing of largescale AGV systems in real-world material handling environments. In theory, all operations can be completed just as planned with no disturbances. However, many deviations may arise during operation, such as vehicle collisions and deadlocks. In real-world applications these deviations cannot be disregarded. The main contribution of this research project is the integration of both a centralized and decentralized approach into one control framework to simultaneously solve the scheduling, collision-free routing, and simulation of dispatching activities for large-scale dynamic problem instances.

Institution: KU Leuven

1. GENERAL INFORMATION

Name applicant

Sander Teck

FWO Project Number & Title

Project title: The development of an effective control framework for large-scale autonomous robot applications.

Project number: 1S97322N

Affiliation

The Centre of Industrial Management / Traffic and Infrastructure (KU Leuven)

2. DATA DESCRIPTION

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

During this project data will be used to generate problem instances that closely relate to real-life operational settings in the warehousing sector. Therefore, they will contain artificial data modeled after real data (order arrival patterns, number of available SKUs and storage distributions) from at least one company in the warehousing industry, and entirely fictional data that is used to generate simplified test cases. This allows for the creation of an extensive list of test cases that will be used to validate the developed algorithms in this research.

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).

The goal of this project is to develop an effective scheduling framework for complex planning and routing problems involving the synchronization of autonomous mobile robots and other resources. The framework will be developed in Python (stored as .py). All work packages after WP1 and not including the final work package (dissemination) will consist of the development of the framework (WP2 – WP6). To validate both the efficiency and effectiveness of this framework various test cases have to be generated. To create these cases data will be generated and collected to make the test cases closely represent the real-world conditions (WP1). For instance, the data will contain order arrival patterns, data on the automated warehouse settings, and technical characteristics of the autonomous mobile robots (AMRs). These datasets are JSON (stored as .json) formatted files since these can efficiently store simple data structures and objects in JavaScript Object Notation (JSON) format. The outputs (e.g. simulation results) from the framework, which is used to analyze the generated schedules' performance, is stored either in CSV files (stored as .csv) or Excel files (stored as .xlsx). Concerning the total volume of the complete dataset at the final phase of this project is estimated at 4 GB.

Some datasets are artificially generated. From company data we extract the order arrival patterns, order/order-line distribution, number of SKUs, AMRs, picking stations and storage policy in real-world operational warehouses and use this data to generate the previously mentioned artificial problem instances. These generated problem instances will be made publicly available at the website of our research group (https://www.mech.kuleuven.be/en/cib/op) since this would benefit the operations research community and can be used as benchmark instances. Important to note is that only those problem instances will be published that comply with the non-disclosure agreements made with the companies that I collaborate with.

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, in this research the focus is solely on developing software for the algorithmic scheduling framework. To validate the framework datasets will be collected and generated, however, they will not contain personal data.

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, there are no ethical issues in this study, all experiments are performed in a simulation environments and do not concern experiments on living beings. For this project, dual use is not expected to result in any ethical issues as well.

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, the research data itself will not be used for potential tech transfer and valorisation. However, the software developed during this project can be of interest to third parties and has a valorisation potential. Therefore, there is a possibility that the software can be licensed under a commercial license. We do not foresee any IP restrictions since it is very difficult to properly protect algorithms with IP.

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?

Yes, there are third party agreements concerning some raw data, from which we generate artificial instances, that restrict dissemination or exploitation. This data relates to more in depth insights on order arrivals, order/order-line distributions, storage policies, etc. However, the sensitive parts of this data is filtered out and anonymized before publishing the datasets and do not limit the ability of publishing any further.

4. DOCUMENTATION AND METADATA

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

A. The source files generated during this research project will be documented in on a GitLab server accompanied with comprehensive and descriptive build and launch script in a readme file format to facilitate the reusability of the software.

B. The developed algorithms are explained in separate academic papers, with pseudocode notation (meta)heuristics are described in such a way that academic researchers are able to reproduce them. Moreover, mathematical modelling notations will be included in several of these papers providing a theoretical view of the handled problems.

C. The generated/collected datasets used for the instance generation are stored in .csv files and will be made publicly available in a version that complies with the NDA agreement of third party companies that provide more in depth information concerning the interested automated warehouse systems. These files will be structured in a self-explanatory manner, with additional notes for clarification.

D. The simulation results obtained from the developed simulation tool will be stored in .csv files, yet again in a very comprehensive manner which allows for easy reusability of the obtained data. Moreover, the results will also be discussed in detail in separate academic reports.

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, there will be no metadata standard used. Of course, for the source code readme files will be created to enhance the reusability of the developed scripts. However, common best practices will be used, which ensures that these files are structured consistently and in a comprehensive manner. For instance, the overall software will be structured into various directories. In each of these directories readme files are added to describe its functionality and how to use it.

5. DATA STORAGE AND BACKUP DURING THE FWO PROJECT

Where will the data be stored?

The developed software will be stored on a GitLab server and is version controlled. The remainder of the data is stored in cloud-based storages, which synchronize with the local desktop storage. This is provided by the KU Leuven Enterprise Box and also provides a version control. The ICTS and the local IT department from the KU Leuven Mechanical Engineering group can support this project if some more specific storage solutions are required.

How is backup of the data provided?

The data will be stored on the university's central servers with automatic daily back-up procedures (OneDrive), with a cloud capacity of 2TB, whereof the access is limited to the applicant and the supervisor. The size of the problem instances are limited and will not require excessive hardware requirements. Source code for the optimization framework will be developed under version control (Sourcetree) with back-up on GitLab servers. So in case of data loss due to hardware failures, the software and datasets can be easily regained.

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 two chosen storage locations (GitLab and OneDrive) have both sufficient storage capacity to store all the generated and collected datasets as well as the source code.

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

The data storage costs of GitLab and OneDrive are covered by the ICT department of the mechanical engineering department. However, if the storage capacity would not suffice for some unforeseen reason, the yearly operating allowance provided by the FWO are more than enough to cover the extra costs of additional storage.

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

Both the datasets and the source code will be stored on the KU Leuven servers (GitLab or OneDrive) where strict authentication is required to access them. Whenever, parts of the study are published, it will also be discussed with the promotor of this study whether certain sets of instance or data are to be made publicly available. This with the intention to contribute to the operational research scientific community.

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, described in Section 4, will be preserved on the KU Leuven data storage facilities for a 10 year period after the end of the project, conform with the KU Leuven RDM policy.

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

The source code developed during the project will remain on the GitLab servers as well as after the project has concluded. The other data (see Section 4) will be stored on the OneDrive of the promotor of this project which includes an automatic daily back-up procedure.

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

Since the promotor will maintain it on his OneDrive the cost of long term storing the data on the KU Leuven ICTS servers is zero each year. If for some reason the storage capacity would not suffice, which we highly doubt, this cost of the long term data preservation will be covered by the operational fund received from the granted project and will be paid upfront to guarantee data preservation for the full 5 year period and even longer.

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)?

No, the data that will be used to generate artificial problem instances will by anonymized and filtered of all the information that would pose a problem for external organizations. Therefore, there are no factors restricting or preventing the sharing of the data.

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

All the data concerning the problem instances will be freely accessible by the research community. The source code of the developed algorithmic framework will only be available to those who will solely use it for educational or research purposes. The raw simulation data will not be publicly published since on its own it does not provide a large added value to the research community, the added value of these simulation results are provided in the discussion in the academic reports.

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

Access to both the databases and associated source code of the algorithmic frameworks generated under this project will be available for educational, research, and non-profit purposes. This access will be provided using either using an Open Access repository, this is suitable for the instance data for example. The source code for the developed algorithmic frameworks can be requested by mail by other researchers in the operational research community.

When will the data be made available?

This data will be made available as soon as the findings of the project have been published. However, if some parts of the software have a strong valorisation potential we reserve the rights to not publish it and consider (commercially) licensing it.

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

The software will be released under a permissive license. For instance, under Creative Commons Attribution Non Commercial Share Alike 4.0 International. This license allows for external parties to work and build upon the developed software with a non-commercial intent, as long as they rightfully credit the work and license it under the same terms. Unless we deem the valorisation potential to be sufficient.

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

Since we are not generating extensive amount of data, we consider the cost of sharing the data to be marginal. However, if some costs do arise these will be covered by the granted operational budget of this project.

8. RESPONSIBILITIES

Who will be responsible for data documentation & metadata?

The applicant (Sander Teck) for the FWO strategic basic mandate will be responsible for both the data documentation and metadata.

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

The applicant (Sander Teck) for the FWO strategic basic mandate will be responsible for both the data storage and back up during the project.

Who will be responsible for ensuring data preservation and reuse ?

The applicant (Sander Teck) for the FWO strategic basic mandate and the promotor (Prof. Dr. Reginald Dewil) will be responsible for both the data preservation and reuse. The full responsibility lies with the applicant during the project and after the project has concluded this responsibility will be (partially) transferred to the promotor of this research.

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

The applicant (Sander Teck) for the FWO strategic basic mandate will be responsible for updating & implementing this DMP.