DMP title

Project Name My plan (FWO DMP) - DMP title
Principal Investigator / Researcher Panos Patrinos
Institution KU Leuven

1. General Information

Name applicant

Panagiotis Patrinos

Masoud Ahookhosh

FWO Project Number & Title

G081222N Taming Nonconvexity in Structured Low-Rank Optimization

Affiliation

- KU Leuven
- Universiteit Antwerpen

KU Leuven: Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics

U Antwerpen: Department of Mathematics, Applied Mathematics Group

2. Data description

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

- Generate new 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).

All data used to support research findings or to validate research results will be generated by the supervisors and co-workers. It will mainly consist of documents in common electronic document formats (reports, papers), analytical and numerical simulation codes in software specific file formats (Matlab m-files, Julia, Python and C code), raw numerical data and multimedia files. The raw data will be generated as output from the simulation codes and will be exported in a variety of standard formats that can be imported in other software packages for post-processing. Metadata about the data sets with specific information about the numerical runs will be stored in associated text files

Type of data	Format	Volume	How created
publications	.pdf	1-10MB	the pdf files generated by latex for submission to journals and conferences
Algorithm iterates, and all other data outputs used for plotting	.dat	1-10 GB	Collected algorithm iterates on benchmark datasets using Matlab, Python, Julia or C++
plots, multimedia files	.pdf, .png	500 MB	plots and videos generated from the iterates
numerical codes and demos	.jl, .py, .mat	0.1-10 MB	Python, Julia, MATLAB, C++ code for solving structured low-rank optimization problems
datasets used in simulations	.csv	1-10 GB	publicly available datasets used in experiments

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

Privacy Registry Reference:
Short description of the kind of personal data that will be used:

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. README files with information about the project, installation instructions, a short example or tutorial, license information, citation information, etc.
2. Open source software licence
3. Jupyter notebooks as a tool for data documentation

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

5. Data storage and backup during the FWO project

Where will the data be stored?

The data will be stored on our STADIUS DATASET Server, which is our research unit central storage server. Copies can be made and kept on personal devices.

To allow for efficient collaboration with researchers from other research groups, the data will additionally be stored on GitHub (subject to the Github regulations).

How is backup of the data provided?

The data will be kept for future use on the ESAT servers. Data in the ESAT servers is backed up daily. The data is also replicated to an off-site storage system housed in the ICTS data center. The compressed data will be stored for at least 5 years after the research project is concluded.

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 STADIUS DATASET Server has a total capacity of 14.88 TB, with a possibility of extention. The capacity of the dataset server is monitored daily by the ESAT system admins.

What are the expected costs for data storage and back up during the project? How

will these costs be covered?

No costs or low costs are expected related to data storage and back up. Potential costs will be covered from the operational costs.

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

The access of the data on the STADIUS DATASET Server is regulated by an access control list (ACL) that grants:

- read-write access to the project owner and the FWO fellow - read-only access to specific users

The ACL is managed by the project owner (Panagiotis Patrinos). Client computers can access the data using:

- SMB2 (or higher) from specific IP ranges
- NFSv4 from specific (IT managed) systems

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 research data detailed previously will be retained for at least 5 years, conform the FWO data preservation policy.

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

The data will be stored on the university's central servers (with automatic back-up procedures) for at least 10 years, conform the KU Leuven RDM policy.

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

The storage facilities of the research division are available at no cost to the project.

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

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

In an Open Access repository

1. The research articles submitted to journals and conference proceedings will be made publicly available on ArXiv e-prints archive.

2. The source code will be released publicly on GitHub under the MIT license, along with the corresponding datasets.

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

- In an Open Access repository

When will the data be made available?

- Upon publication of the research results

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

1. The research articles submitted to journals and conference proceedings will be made publicly available on ArXiv e-prints archive under the perpetual, non-exclusive license. Therefore, it will be available to anyone for any purpose, provided that they give appropriate credit to the creators.

2. The source code will be released publicly on GitHub under the MIT license, along with the corresponding datasets. Hence, it will be available for everyone to use, change, and distribute the software, only requiring preservation of copyright and license notices.

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

As free services like GitHub and ArXiv are used to distribute the research data, there are no expected costs.

8. Responsibilities

Who will be responsible for data documentation & metadata?

The PIs (Panagiotis Patrinos and Masoud Ahookhosh) will be responsible for data documentation and metadata.

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

The PIs (Panagiotis Patrinos and Masoud Ahookhosh) will be responsible for data storage & back up during the project.

Who will be responsible for ensuring data preservation and reuse ?

The PIs (Panagiotis Patrinos and Masoud Ahookhosh) will be responsible for ensuring data preservation and reuse.

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

The PIs bear the end responsibility of updating & implementing this DMP.