Plan Overview

A Data Management Plan created using DMPonline

Title: 18651 ADAPTOR: Autonomous Distribution Architecture on Progressing Topologies and Optimization of Resources

Creator: Georgios Exarchakos

Principal Investigator: George Exarchakos

Affiliation: Other

Funder: Netherlands Organisation for Scientific Research (NWO)

Template: Data Management Plan NWO (September 2020)

ORCID iD: 0000-0002-5773-0720

Project abstract:

Computing resources available at or nearby a Cyber-physical system have much slower upgrade cycles than the algorithms they serve. This makes maintenance very expensive as physical access by experts becomes necessary for upgrades or complete replacements. ADAPTOR extends the autonomy of these systems by building an intelligent fog/edge solution able to aggregate all resources of interconnected devices into a single distributed pool and assign tasks to it. At the core of the solution is an adaptive distributed scheduler of tasks to resources able to detect emerging patterns and serve unforeseen applications, hardware performance and usage patterns. The distributed scheduler tries to accurately predict the availability of resources at the moment of resource commissioning to maximize resource utilization and task throughput. Both our diverse use cases, electron-microscopy and space-based radio telescope swarms, bring challenges on time-sensitivity, tight resource constraints and scalability.

ID: 75091

Start date: 09-01-2021

End date: 08-03-2025

Last modified: 09-04-2021

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

18651 ADAPTOR: Autonomous Distribution Architecture on Progressing Topologies and Optimization of Resources

General Information

Name applicant and project number

Project Leader: Georgios Exarchakos
Project Number: 18651

Name of data management support staff consulted during the preparation of this plan and date of consultation.

Sil van Lieshout, Data Steward – Research Data Management, Data Management & Library, email: h.j.v.lieshout@tue.nl
Consultation date: **07**/**0**4/2021

1. What data will be collected or produced, and what existing data will be re-used?

1.1 Will you re-use existing data for this research?

If yes: explain which existing data you will re-use and under which terms of use.

- Yes

ASTRON will provide open access datasets from recorded cosmos signals from various orbiting satellites. This data are open already for use by any researcher around the world. It includes mostly experimental measurements and software for pre-processing.
TU/e has a verbal agreement with Thermo Fisher Scientific on necessary data from their systems. Thermo Fisher Scientific will provide performance data from their electron microscopes computing units and clouds. This data is currently held in house and can be available for this project. It includes mostly experimental measurements and software for pre-processing.

1.2 If new data will be produced: describe the data you expect your research will generate and the format and volumes to be collected or produced.

The project is not targeting data collection or generation. Yet, any experimentation data and software used to assess the performance of the developed systems will be made openly accessible together with the publications associated with them. The data will be mostly numeric data and software. The format of the numeric data will be tabular in CSV structure. The software will be made available via open access repositories like github/gitlab.

1.3. How much data storage will your project require in total?

- 0 – 10 GB

Any data produced in this project will eventually accompany scientific publications. The data and software will therefore become openly available online. They will stay available in the 4TU.ResearchData repositories and cited in relevant publications.

2. What metadata and documentation will accompany the data?

2.1 Indicate what documentation will accompany the data.

The actual publications associated to the data as well as README files explaining the structure and interpretation of the data will be inseparable information bundled with the actual data. Procedures implemented by certain software to use that data will accompany the software published in open software repositories. Necessary linking between that software repositories and data bundles will be present in INSTALL files included in those repositories. Each software repository will also include a wiki documentation with more detailed explanation of the code designed and used as well as the exact procedure of regenerating the results in publications.

2.2 Indicate which metadata will be provided to help others identify and discover the data.

During the research, data and software will be stored on git facility provided by the Information Management and Services department of the TU/e. They will remain available to the research group by using a clear folder structure and file names.

This project will use ISA-TAB metadata standard to describe the data before deposited. The exact ISA-TAB template we will use, will be decided later in due course.

Following the completion of the project, relevant data/software (i.e. published and unpublished work) are stored for the long term (10 years) with controlled access on secure network drives available at the departments. Daily back-ups are made on this storage facility. Relevant data that are not subject to access restrictions (e.g., confidentiality, competitive concerns and IP, personal data and otherwise sensitive data) will be disseminated via a data repository. In our case the data repository of choice will likely be 4TU.Research Data (https://data.4tu.nl). Data published in 4TU.Research Data are available open access and data availability is guided by the FAIR principles with sufficient metadata. DOIs in the publications will be linked to the datasets to make them easily findable

3. How will data and metadata be stored and backed up during the research?

3.1 Describe where the data and metadata will be stored and backed up during the project.

- Institution networked research storage

Given that the project data will be pulled from collaborators i.e. ASTRON and Thermo Fisher Scientific, the project will use Research Drive (RDRIVE) of Surfnet to manage the data sources.

TU/e, via SURF, has made available a data and metadata storage service per project: the Research Drive - https://wiki.surfnet.nl/display/RDRIVE/

This is the location the project will use to organize and share all data used and produced. This data storage and sharing environment allows for use of multiple external linked storage spaces. The project will not rely on external storage spaces though. All data generated and used in publications will be maintained exclusively on SURF Research Drive.

RDRIVE will link data from ASTRON, Thermo Fisher Scientific and TU/e git repos used by the project.

3.2 How will data security and protection of sensitive data be taken care of during the research?

- Not applicable (no sensitive data)

4. How will you handle issues regarding the processing of personal information and intellectual property rights and ownership?

4.1 Will you process and/or store personal data during your project?

If yes, how will compliance with legislation and (institutional) regulation on personal data be ensured?

- No

4.2 How will ownership of the data and intellectual property rights to the data be managed?

The principal investigator is the sole owner and responsible of the data. The data will be opened to anyone, for read access only, once the related publications are accepted by the publisher of the work. Before that point, the internal and external researchers directly involved in the project (TU/e, ASTRON and Thermo Fisher Scientific) will be given access to read and write. Once the articles are published the related data and software will be made publicly available with read access only so that any reader can download and re-use it.

This covers protection of intellectual property rights as the data will be only openly available to anyone at the same time the relevant publications are also open.

5. How and when will data be shared and preserved for the long term?

5.1 How will data be selected for long-term preservation?

- All data resulting from the project will be preserved for at least 10 years

No data will be needed to be destroyed. Any data stored in the SURF Research Drive will stay open for at least 10 years after the project end.

5.2 Are there any (legal, IP, privacy related, security related) reasons to restrict access to the data once made publicly available, to limit which data will be made publicly available, or to not make part of the data publicly available?

If yes, please explain.

- No

5.3 What data will be made available for re-use?

- All data resulting from the project will be made available

Any data stored in the 4TU.ResearchData repositories will stay open and available to all for re-use for at least 10 years after the project end.

Surf ResearchDrive will be used for data source management and backup during the execution of the project.

TU/e IMS Git repositories will be used for the day-to-day development process of software.

5.4 When will the data be available for re-use, and for how long will the data be available?

- Data available as soon as article is published

5.5 In which repository will the data be archived and made available for re-use, and under which license?

4TU.ResearchData drive will be the repository for archive, too. The data will be released with the CC-BY creative commons license giving the maximum freedom to anyone who is interested to use the published data. Similarly, the software will be released under MIT license to allow any use of the software.

5.6 Describe your strategy for publishing the analysis software that will be generated in this project.

As mentioned above, the software will be released publicly under MIT license in git versioning repository e.g. github or gitlab.
The software in those repositories might get updated and expanded by several other researchers over the years. Yet, the software will be organized in releases. These releases will be registered to the 4TU software registry (data.4tu.nl).
The FigShare functionality of the 4TU registry will generate a globally unique identifier (DOI) per release archived in the registry. This DOI will make each release citable.

6. Data management costs

6.1 What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

The institutional network services for data storage, back-ups and archiving of data are already available and there will be no extra costs for this, since these costs are part of the budget of the department.
Any data and software publicly released in this project will be curated and shared by the research article authors based on the plan above. The chosen repositories and data storage places do not entail extra costs for their use.
Making data FAIR (assigning DOI, documenting data, assigning metadata to the data, cleaning/tidying the data, etc.) is poart of the ongoing research activities of the involved researchers.