Assuring the safety of autonomous systems through human-centered executable assurance cases

A Data Management Plan created using DMPonline.be

Creator: Laure Buysse

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Grant number / URL: 1S17123N

ID: 198793

Start date: 01-11-2022

End date: 31-10-2026

Project abstract:

While autonomous systems offer tremendous possibilities, they also come with major safety challenges. After all, human safety is central to every autonomous system. Existing safety assurance approaches and standards have been developed primarily for systems where a human can take over in the case of emergency and do not extend to autonomous systems. The combination of executable safety assurance cases and digital twins has tremendous potential as a solid safety framework for autonomous systems. However, the transition from classic static-safety assurance during design-time to dynamic safety guaranteed by the system itself without further human intervention is a big step to take. Therefore, this PhD proposal will focus on the concept of human-centered executable safety cases as an important intermediate step. More specifically, we will investigate the following research hypothesis: can we enable the safety engineer to dynamically re-evaluate the safety claims, arguments and assumptions underlying the safety assurance case and select the appropriate course of action by combining the concepts of executable safety assurance cases and digital twins. This research will create fundamental knowledge on dynamic risk management, necessary to assure the safety of modern high-tech, software-driven autonomous systems. Additionally, the results of this research supports the transition towards increased (optimal) use of smart devices and safe intelligent transport systems.

Last modified: 25-04-2023

Assuring the safety of autonomous systems through human-centered executable assurance cases
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.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	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, .cvs,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Safety Cases Dataset	Dataset containing (partial) safety cases of all studied system	Generate new data	Digital	Observational	.pdf (figures) .xml (metadata)	< 100MB	
Digital Twin Model AMR	A digital twin model (made in Gazebo) of an Autonomous Mobile Robot	Generate new data	Digital	Software	Gazebo files	< 1GB	
Dynamic Safety Case Ruleset	Set of selection rules which forms the basis for the transfer of static safety cases into runtime	Generate new data	Digital	Observational	.py .docx / .txt (documentation)	< 100 MB	
HESC Tooling	Software application demonstrating the theory around human-centered execurable safety cases (HESC)	Generate new data and reuse exiting data	Digital	Software	.py .md / .txt (documentation)	< 100 MB	
AMR Dataset	Dataset containing data from an autonomous mobile robot (at KU Leuven Bruges Campus) to use for demonstration (within the HESC and Safety prediction tooling) and aid the devepmet of the Dynamic Safety Case Rulest and Safety Prediciton algorithms within the Safety prediction tooling	Generate new data	Digital	Observational	.json or .csv .docx / .txt (docuentation)	< 100GB	
Safety prediction tooling	Set of safety prediction algorithms (some integrated with the digital twin model) to allow for short-term and long-term prediction of safety based upon runtime safety case data	Generate new data	Digital	Software	.py .mdf / .txt (documentation)	< 1GB	

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:

To aid with the formal monitoring component of the HESC tooling, th C++ / python library "Reelay Monitors" by Dogan Ulus will be used. The library can be found on github: https://github.com/doganulus/reelay (Ulus, Doğan. (2019). Online Monitoring of Metric Temporal Logic using Sequential Networks.).

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

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

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.

- Yes

The HESC tooling, which implements the PhD concept, has the potential for commercial exploitation. The contents of this library / these scripts can be applied to complex and / or autonomous (inudstrial) systems (such as autonomous mobile robots in a warehouse) to improve general safety, to aid the discovery of safety gaps or to help with data generation.

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

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

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).

Measurement datasets will be accompanied by linked guidance documents.

All safety case related material (Safety case dataset - possible demo's for tooling) will be saved using the appropriate industry standards, e.g. the SACM standard or the GSN community standard.

Tooling (HESC, DT and safety prediction) is thoroughly explained with in-script annotation and additional guidance documents. Furthermore, demo scripts will be created to extensively demonstrate functions and usage to assistuture reuse.

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.

- Yes

The "Safety Cases Dataset" will use either of two major industry standards (depending on the type of safety case):
- the "Structured Assurance Case Metamodel" (SACM) standard by the "Object Management Group"
- the GSN community standard
The "AMR Dataset" will be stored using a standardised JSON format as exported from InfluxDB.

3. Data storage & back-up during the research project

Where will the data be stored?

Active tooling (incl. demos) are uploaded and stored using Gitlab.
Research data which is currently not shared (such as the 'Safety case dataset') is stored on a personal
OneDrive for Business cloud storage provided by KU Leuven.

How will the data be backed up?

Local data and files in general are backed up using OneDrive for business' file synchronisation. Alternatively, the data stored in GitLab is additionaly stored on the researchers personal hard drive.

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

With the estimations for each dataset, it is clear that 1 TB of storage should be more than sufficient for the project. This means the OneDrive for Bussiness storage, capable of storing 2 TB should more than suffice. For tooling, each gitlab repository has a maximum of 10 GB which should also suffice, additionally the personal hard drive of 1 TB is also sufficient for backup storing.

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

Any acces to the files on OneDrive for business will be given through the usage links with strict access only to the intended receiver.

The GitLab repositories are moderated by myself. Thus any acces has to be given by msyelf, which can be done with various rights to any additional user as needed (read-only, read-write, admin ....).

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

No costs are expected as of now. The 10GB free storage for the GitLab repositories should suffice as well as the 2TB OneDrive for business storage provided by KU Leuven.

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 stored in the KU Leuven Research Data Repository (RDR) for a minimum of ten years in accordance with the KU Leuven RDM Policy. Data will be linked to corresponding publications where applicable.

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

Data will be stored in the KU Leuven Research Data Repository (RDR) for a minimum of ten years. Data will be linked to corresponding publications where applicable.

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

No additionals costs for data preservation are expected as the KUL RDR provies 50GB long-term storage by default.

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 a restricted access repository (after approval, institutional access only, …)

The dataset and tooling will be / is reused in guided student projects. This reuse mainly concerns the "HESC tooling" and "Digital twin model".

If access is restricted, please specify who will be able to access the data and under what conditions.

Access to the GitLab tooling repositories is authorized by myself. This will be given to the students who require the tools and supervisors. Teh type of access (read-only, read-write, admin) will be determined based upon the requirements project requirements.

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.

Data will be made available to supervisors through a shared link for the OneDrive Folders or through the GitLab repositories.

When will the data be made available?

The "Safety Cases Dataset" has been made available in OneDrive since 23/11/2023.
Access to all tooling and other datasets will be made available as the reserach progresses.

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

At the end of the project, data will be made publicly available in RDR using Public Domain Mark (PD) for datasets and the Mozilla Public License 2.0 for the created software.

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?

The costs for data sharing in KUL RDR are free of charge up to 50GB, larger datafiles can be covered using the FWO bench fee if need be.

6. Responsibilities

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

Laure Buysse. Those who upload new data are initially responsible for providing data documentation of their data.

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

Laure Buysse

Who will manage data preservation and sharing?

Laure Buysse

Who will update and implement this DMP?

Laure Buysse