European Training Network on

Safer Autonomous Systems (SAS)

D5.5 – Data Management Plan (v1.0)

Contents

1. Executive Summary

This document details the MSCA ETN SAS Data Management Plan. It describes the data management life cycle for the data to be collected, processed and/or generated by the MSCA ETN SAS project. This document adopts the Horizon 2020 FAIR DMP Template. The H2020 FAIR DMP template is a set of questions that are answered with a level appropriate to the project. Detailed answers to all the questions are not yet provided in the first version of the DMP. Rather, the DMP is a living document in which information is made available on a finer level of granularity through updates as the implementation of the project progress and when significant changes occur.

2. Data Summary

MSCA ETN SAS (Safer Autonomous Systems) brings together early stage researchers and experienced specialists from key players in academia and industry across Europe covering different scientific disciplines and industrial stakeholders from a broad range of backgrounds to optimally address the challenges ahead. The researchers will be trained in innovative PhD topics as well as receiving specific theoretical and practical training in the field of safe autonomous system engineering, tackling aspects of (functional) safety, the design and implementation of autonomous systems, and verification and validation of systems.

The SAS Consortium is formed by 7 beneficiaries combining leading education institutes, top research institutions and leading companies as well as 10 partner organisations spanning multiple disciplines such as, automotive, agriculture, avionics, medical and maritime, to assist in the dissemination and public engagement of the SAS results. Furthermore, this consortium will provide dedicated training to enhance the entrepreneurial mind-set of the ESRs. The researchers will profit from top scientific research guidance in combination with highly relevant industrial supervision. Moreover, the industry will gain from the specific training of the researchers.

In achieving the above, SAS will collect and generate:

- Technical data (including test and simulation data) within the technical work packages WP 1, WP 2 and WP 3.
- Data for training, exploitation, dissemination and communication, and management purposes in WP 4, WP 5 and WP 6.

For the MSCA ETN SAS Project, the Consortium has chosen to opt-out of the Open Research Open Data Pilot (ORD pilot). The reason for opting out is to allow the protection of results (e.g. patenting) as stated in the proposal submission form. However, researchers and their supervisors might make their data open on a voluntary basis. In such case, that data will be made available according to the FAIR principles and the data management plan will be updated (Section 3) to reflect this.

The tables below indicate the kind of data generated and collected within the different SAS tasks (according to the best knowledge of the SAS Consortium at the moment of writing).

ESR	Supervisor	WP/Task	Data type	Data format & estimated size	Origin of data	Data Dissemination level & platform	Data utility
WP1 Designing inherently safe autonomous systems							
1	LAAS	1.1 Development of a generic framework to monitor and handle safety of autonomous systems during run-time	Simulation results	Safety monitor prototype (program)	From researchers from self-defined cases	Public	Academic and industrial researchers and engineers working in this field
2	Fraunhofer	1.2 Development of an adaptive platform for resilient autonomous systems based on a MAPE-K cycle	Analytical results	Word/Excel documents	From researchers from self-defined cases	Via the specific deliverable and through public dissemination	Other scientists in the field
3	Fraunhofer	1.3 Dynamic safety handling of autonomous systems-of-systems with run-time safety contracts	Analytical results	Word/Excel documents	From researchers from self-defined cases	Via the specific deliverable and through public dissemination	Other scientists in the field
4	KU Leuven	1.4 Creating Software Design Guidelines and Testing Specifications for Non-Functional Requirements in Safety-critical Autonomous Systems	Source code, simulation data	Text files, Excel files (or simular), figures, simulation projects,... Size: order 1 TByte	From researchers from self-defined cases	Via the specific deliverable and through public dissemination	Academic and industrial researchers and engineers working in this field
5	KU Leuven	1.5 Making Connectivity Work Reliably in a diverse Range of Environments	Simulation results, measurement data	Text files, Excel files (or simular), figures, simulation projects,... Size: order 1 TByte	From researchers from self-defined cases	Via the specific deliverable and through public dissemination	Academic and industrial researchers and engineers working in this field

ESR	Supervisor	WP/Task	Data type	Data format & estimated size	Origin of data	Data Dissemination level & platform	Data utility
WP2 Providing evidence for autonomous systems							
6	LAAS	2.1 Virtual worlds generation for testing autonomous robots in simulation	Simulation results	Program and execution traces	From researchers from self-defined cases	Public	Academic and industrial researchers and engineers working in this field
7	UoYork	2.2 Rigorous Design and Evaluation of Situation Coverage Testing for Autonomous Vehicles	Code Simulation files, situation coverage data files, final result files	Code — C or Java or Python Input and output data files — unknown at this time Size — no way to estimate now, probably multiple GB	From researchers from self-defined cases, virtual environments	UOY website, GitHub etc. Free download with clear instructions	Allows our experiments to be reproduced, expanded, improved.
8	Bosch	2.3 Model-based System Analysis Techniques to determine propagation paths of functional insufficiencies in software-intensive systems	Analytical results	To be determined	From researchers from self-defined cases	To be determined	Other scientists in the field
9	KU Leuven	2.4 Model-based System Analysis of the Robustness of Autonomous Systems against ElectroMagnetic Interference	Simulation results, measurement data	Text files, Excel files (or simular), figures, simulation projects,... Size: order 1 TByte	From researchers from self-defined cases	Via the specific deliverable and through public dissemination	Academic and industrial researchers and engineers working in this field

ESR	Supervisor	WP/Task	Data type	Data format & estimated size	Origin of data	Data Dissemination level & platform	Data utility
WP3 Providing assurance strategies							
10	UoYork	3.1 From static assurance cases at design-time to executable assurance cases at run-time	Numerical input, data from secondment organisation, code for formal verification	Program size to be determined	From researcher or the use of open source code	Publicly on Github with instructions to run and via the specific deliverable	Other scientists in the field
11	UoYork	3.2 Assurance case structures for machine learning in the decision making of highly autonomous systems	Open Source Image - Data for Example (http://benchmark.ini.rub.de/) Modifications to and scripts for Open Source simulators like Carla, SelfDriving by Udacity	Images and Videos, size based on availability Python code	http://benchmark.ini.rub.de/ http://carla.org/ https://github.com/udacity/self-driving-car-sim + own modifications and scripts	Original data hosts as much as possible on Github and on request	Reproducability of results
12	RH Marine	3.3 Assuring autonomous sailing from A to B while minimizing operational costs	Simulation results	To be determined	From researchers from self-defined cases	To be determined	To be determined
13	UoYork	3.4 Safety assurance for Clinical Conversational Bots	To be determined	To be determined	To be determined	To be determined	To be determined
14	Horiba-MIRA	3.5 Dependability Assurance for Vehicle Autonomy	Framework for assurance cases. Classifications and ranking schemes. Graphical process models.	Text files, Excel files (or simular), figures, simulation projects,... Size: order 1 TByte	From researchers from self-defined cases. Anonymized real-world data extracted from client projects.	Via the specific deliverable and through public dissemination	Academic and industrial researchers and engineers working in this field
15	KU Leuven	3.6 Between Safety and Liability: Towards a Liability Allocation Framework for Safe Autonomous Systems	Qualitative data. Most of the data input and output will not be numerical. No sensitive data is expected to be used during the course of the study (except for confidential notes or interviews provided for research purposes).	Text files, Excel files (or simular), figures, simulation projects,... Size: order 1 TByte	- existing legislation - case law - reports and studies - online legal journals and scholarly research papers (online portals) - academic books - interviews or confidentional notes	In principle, the research deliverables (and some publications) will be made available to the public. However, research input and prepatory material will not be published, unless necessary. Moreover, confidential interviews, notes or reading material will be subject to limited or no public release.	Multiple uses: collected data will enable the researcher to assess the legal frameworks. Target audience of the output data: academic and industrial researchers, legal scholars and policymakers.
WP4 Training							
-	All	4.1 Individual scientific-technological and complementary skills training	Course material	Word documents, PPTs and PDFs	From researchers and lecturers	Confidential training information for the consortium only	Consortium, training progress
-	All	4.2 Network-wide training (including scientific and complementary skills modules)	Course material	Word documents, PPTs and PDFs	From researchers and lecturers	Confidential training information for the consortium only	Consortium, training progress
WP5 Exploitation, Dissemination and Communication							
-	All	5.1 Internal dissemination	Intranet platform, tele- and videoconferencing, NWEs	Recordings, HTML-pages and PDFs	From researchers, supervisors and MST	Consortium level via the Intranet platform	Consortium
-	All	5.2 External dissemination, outreach and communication	Website, Social Media and Publications	HTML-pages and PDFs	From researchers, supervisors and MST	Public via the website and publicly available publications	Public, policy makers and research managers
WP6 Management							
-	KU Leuven	6.1 Project start-up	Presentations	PPTs and PDFs	From Supervisors and MST	Consortium level via the Intranet platform	Consortium

3. FAIR Data – Findable, Accessible, Interoperable and Re-usable

An overview of all data that will be collected or generated within the framework of MSCA ETN SAS Project, including their intended dissemination levels is presented in the table under **Data Summary**.

This section of the DMP is filled out based on the data provided its owners. This section will eventually grow during the project as this DMP is a living document.

3.1 Making data findable, including provisions of metadata

WP Task Deliverable	Data type	Data format	Unique identifiers	Data naming conventions	Metadata keywords	Upload to a searchable resource

3.2 Making data openly accessible

WP Task Deliverable	Data type	Data format	Access: method	Access: location	License type	Documentation provided

3.3 Making data interoperable

WP Task Deliverable	Data type	Data format	Metadata keywords	Documentation provided	Reuse of data

3.4 Increase data re-use (through clarifying licenses)

WP Task Deliverable	Data type	Data format	License type	Data Quality Assurance	Retention period	Embargo period

4. Allocation of Resources

The costs for collecting, processing, packaging and making the data available via the ETN SAS website and/or other repositories are covered by the MSCA ETN SAS Grant Agreement.

Also costs to possible Open Access journal publications will be covered by the MSCA ETN SAS Grant Agreement.

5. Data Security & Ethical aspects

Data security is dealt with on local beneficiary level, with all partners being well experienced in H2020 and industrial research projects.

The philosophy of the MSCA ETN SAS Project is based on the premise that 'ethics contributes to responsible research and innovation from the early stages of design to the later stages of application and translation to research'. All research activities within the MSCA ETN SAS Project will be carried out in line with the European Charter & Code for Researchers.

MSCA ETN SAS Deliverable D7.1 details ethical aspects considered.

Deliverable 7.1 will contain the ethical issues regarding the societal impact of autonomous systems describing the SAS project's strategy to respect human rights and values in the development of safer autonomous systems. This deliverable will also contain risk mitigation strategies for the misuse potential of autonomous systems involved in the research.

6. Conclusion

This dedicated deliverable 5.5 DMP outlines the guidelines and strategies for data management of MSCA ETN SAS which will be fine-tuned and extended throughout the lifecycle of the project. This DMP will be updated on a regularly basis as this is pointed out as an Action Item at every 6-monthly MSCA ETN SAS Network Wide Event. Following the H2020 FAIR DMP template, the scope and the purpose of the MSCA ETN SAS datasets are described.