SUNRISE

SAFETY ASSURANCE FRAMEWORK FOR CONNECTED, AUTOMATED MOBILITY SYSTEMS

Deliverable 1.3

Data Management Plan

Project short name
SUNRISE

Project full name
Safety assUraNce fRamework for connected, automated mobIlity SystEms

Horizon Research and Innovation Actions | Project No. 101069573
Call HORIZON-CL5-2021-D6-01

Funded by
the European Union

ccam-sunrise-project.eu/

Dissemination level	Public (PU) - fully open
Work package	WP1: Coordination and Management
Deliverable number	D1.3: Data management plan
Status - Version	V1.1
Submission date	27/02/2023
Keywords	Data, Management, Plan

Quality Control

	Name	Organisation	Date
Peer review 1	Stefan de Vries	IDIADA	25/01/2023
Peer review 2	Jordi Pont	IDIADA	25/01/2023
Peer review 3	Stefan de Vries	IDIADA	27/01/2023

Version history

Version	Date	Author	Summary of changes
0.1	17/01/2023	Jonathan Smith Tudor Dodoiu	First draft
0.2	19/01/2023	Jonathan Smith	Material edits and structural changes
0.3	26/01/2023	Jonathan Smith	Response to IDIADA review. New personal data and GDPR content added.
1.0	30/01/2023	Jonathan Smith	Final version
1.1	27/02/2023	Jonathan Smith	Minor update based on feedback EC

Legal disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND ACRONYMS

Abbreviation	Meaning
CA	Consortium Agreement
CC by-NC	Creative Commons Non-Commercial license
CC by-ND	Creative Commons Attribution NoDerivs
CCAM	Cooperative, Connected, and Automated Mobility
CORDIS	Community Research and Development Information Service
DM	Dissemination Manager
DMP	Data Management Plan
DPO	Data Protection Officer
EC	European Commission
ERTRAC	European Road Transport Research Advisory Council
EU	European Union
FAIR	Findable, accessible, interoperable and reusable
GA	Grant Agreement
GDPR	General Data Protection Regulation
GPS	Global Positioning System
I&EM	Innovation and Exploitation Manager
LM	Liaison Manager
ODD	Operational Domain Design

PoC	Proof of Concept
SDL	Scenario Definition Language
UC	Use Case
V&V	Verification and Validation
WP	Work Package

EXECUTIVE SUMMARY

This Data Management Plan (DMP) builds on the project's contractual requirements to provide guidance and best practice for data management within the SUNRISE project. The DMP places emphasises on three points:

- The need to protect commercially sensitive data
- The aspirations of the project to publish and disseminate its findings wherever possible
- The need to comply with GDPR and/or other relevant personal data laws and requirements

Execution of the DMP is the responsibility of all project members relative to their roles and responsibilities. The Project Coordinator, WP leaders and Task leaders, under the guidance and supervision of the DPO, will provide wide reaching project support to implement the DMP and provide additional support where required.

The DMP describes the FAIR best practice for data management to ensure data is made findable, accessible, interoperable, and reusable (FAIR).

The DMP provides guidance on how to label data, classify data and the processes through which partners should go before making any data publicly available. This includes guidance on how to identify who is responsible for a data set, along with respective roles and responsibilities of project members relative to their ownership of or usage of data.

This Data Management Plan (DMP) will elaborate to define procedures on how to handle personal data to guarantee project participants' fundamental rights and avoid misuse of the project results. The plan complies with EU and international regulations for the management and use of data and will pay particular attention to the General Data Protection Regulation (GDPR), which came into effect in May 2018.

If you are, or think you maybe, dealing with personally identifiable data you will be classed as a 'data controller' under GDPR. Whenever any project member intends on collecting, processing, or using personally identifiable data of any kind, GDPR and other legal compliance requirements must first be considered. This DMP provides:

- The key definitions (Personal data, Data processing, Data subject, Data controller, Data processor)
- The key principals that should be followed (Lawfulness, fairness and transparency, Purpose limitation, Data minimization, Accuracy, Storage limitation, Integrity and confidentiality, Accountability)
- Accessing further guidance

As soon as partners are aware data and findings are of public value, they will begin, in accordance with the guidance in this plan and the terms of the CA and GA, exploring the possibility of making that data public. The DMP provides best practice guidance and processes for making data available within the project and in order that that data be maintained and available beyond the time frame of the project.

Finally, the plan recommends approaches to data security and storage and whether it will be destroyed at the end of the project or archived for further use by the research community. In the latter case, the DMP provides recommendations for future maintenance and access to the data by consortium partners and external parties.

1 INTRODUCTION

The SUNRISE consortium will carefully manage research data generated and collected during the project. The Consortium Agreement (CA) includes provision for data management as agreed between the project partners, and the Grant Agreement (GA) including provisions for data management obligations as agreed with the European Commission. However, good data management also requires best practice guidance and project level processes. This report builds on the projects contractual requirements to provide such guidance and best practice, with emphasises on three points:

- The need to protect commercially sensitive data
- The aspirations of the project to publish and disseminate its findings wherever possible
- The need to comply with GDPR and/ or other relevant personal data laws and requirements

Execution of the DMP is the responsibility of all project members relative to their roles and responsibilities. The Project Coordinator, WP leaders and Task leaders, under the guidance and supervision of the DPO, will provide wide reaching project support to implement the DMP and provide additional support where required.

It should be noted that the DMP is in no way a substitute for the projects legal documents and do not replace their enforcement in any way – if in doubt the legally binding documentation takes precedence.

The DMP describes the FAIR best practice for data management to ensure data is made findable, accessible, interoperable, and reusable (FAIR).

The DMP provides guidance on how to label data, classify data and the processes through which partners should go before making any data publicly available. This includes guidance on how to identify who is responsible for a data set, along with respective roles and responsibilities of project members relative to their ownership of or usage of data.

This Data Management Plan (DMP) will elaborate to define procedures on how to handle personal data to guarantee project participants' fundamental rights and avoid misuse of the project results. The plan complies with EU and international regulations for the management and use of data and will pay particular attention to the General Data Protection Regulation (GDPR), which came into effect in May 2018.

The plan recommends approaches to data security and storage and whether it will be destroyed at the end of the project or archived for further use by the research community. In the latter case, the DMP provides recommendations for future maintenance and access to the data by consortium partners and external parties.

As part of the DMP's production a questionnaire has been shared with all project participants to gather information on the types of data partners foresee generating, managing and sharing.

2 SUNRISE PROJECT SUMMARY

Safety assurance of Cooperative, connected, and automated mobility (CCAM) technologies is a crucial factor for their successful adoption in society, yet it remains to be a significant challenge.

CCAM systems must prove to be reliable in every possible driving scenario, which requires a strong safety argumentation. It is already acknowledged that for higher levels of automation, the validation of these systems by means of real test-drives would be infeasible. In consequence, a carefully designed mixture of physical and virtual testing has emerged as a promising approach, with the virtual part bearing more significant weight in this mixture for cost efficiency reasons. Several worldwide initiatives have started to develop test and assessment methods for automated driving functions. These initiatives have already moved from conventional validation to a scenario-based approach and combine different test instances (physical and virtual testing) to avoid the million-mile issue.

The initiatives mentioned above provide new approaches to CCAM validation, and many expert groups formed by different stakeholders are already working on CCAM systems' testing and quality assurance. Nevertheless, the fact that there is a lack of a common European validation framework and homogeneity regarding validation procedures to ensure safety of these complex systems, hampers the deployment of CCAM solutions. In this landscape, the role of **standards** is paramount in establishing common ground and providing technical guidance. However, standardising the whole pipeline of CCAM validation and assurance is in its infancy, as many of the standards are under development or have been very recently published and still need time to be synchronised and established as common practice.

Scenario databases are another issue tackled by several initiatives and projects, providing silo solutions. A single concrete approach should be used (at least at the European level), dealing with scenarios of any possible variations, including the creation, editing, parameterisation, storing, exporting, importing, etc. in a universally agreed manner.

Furthermore, validation methods and testing procedures still lack appropriate **safety assessment criteria** in order to build a robust safety case. These must be set and be valid for the whole parameter space of scenarios. Another level of complexity is added, due to regional differences in traffic rules, signs, actors, and situations.

Evolving from the achievements obtained in predecessor project HEADSTART and taking other initiatives as a baseline, it becomes necessary to move to the next level in the concrete specification and demonstration of a commonly accepted **Safety Assurance Framework** (**SAF**) for the safety validation of CCAM systems, including a broad portfolio of use cases and comprehensive test and validation tools. This will be done in **SUNRISE**, which stand for **S**afety ass**U**ra**N**ce f**R**amework for connected, automated mob**I**lity **S**yst**E**ms.

The Safety Assurance Framework is the main element to be developed in the SUNRISE project. As the following figure indicates, it takes a central role, fulfilling the needs of different automotive stakeholders that all have their own interests in using it.

The **overall objective** of the SUNRISE project is to accelerate the safe deployment of innovative CCAM technologies and systems for passengers and goods by creating demonstrable and positive impact towards safety, specifically the EU's long-term goal of moving close to zero fatalities and serious injuries by 2050 (Vision Zero), and the resilience of (road) transport systems. The project aims to achieve this by creating and sharing a European federated database framework centralising detailed scenarios for testing of CCAM functions and systems in a multitude of relevant test cases, based on a virtual harmonised simulation environment with standardised, open interfaces and quality-controlled data exchange.

2.1 Objectives

The **specific objectives** of the project are as follows:

Objective 1 - Develop and provide a scalable, future-proof Safety Assurance Framework that fulfils the needs of different bodies (like consumer testing and type approval agencies) and key stakeholders (like vehicle manufacturers and their suppliers) for a continuously evolving number of CCAM use cases, ensuring compliancy with existing standards and enabling the creation of new ones.

Objective 2 - Define safety argumentation principles (providing methods to assess testing fidelity and coverage when different test approaches are used) and introduce a CCAM safety assurance framework handbook with an interactive tool.

Objective 3 - Define comprehensive verification, validation and rating procedures standardized and valid for different stakeholders: 1) Industry and academia developers; 2) Regulatory forums; 3) Consumer testing agencies (NCAPs). Safety metrics utilised as criteria for the assessment procedures, will be defined, showing that the tests are suitable to discriminate between a safe and an unsafe system.

Objective 4 - Implement the mechanisms for scenario allocation in the test toolchain to cover the whole parameter space. The issue of scenario library scaling-up will be also tackled through methods for scenario variations' automatic generation within predefined ODDs and scenario criticality metrics.

Objective 5 - Define and validate a test, assessment and validation toolchain, focusing on the simulation environment and driver's capabilities, in order to promote the standardisation and definition of open interfaces and quality-controlled data exchange.

Objective 6 - Define a harmonised scenario data framework complemented by the definition of its structure, governance mechanisms and scenario description (ontology). This framework will tackle the current complexity due to heterogeneity among existing databases and propose a model of coexistence that maximises completeness and accessibility.

Objective 7 - Demonstrate the framework in a representative set of use cases to prove the robustness, repeatability and versatility of the methodology through a rich simulation kits portfolio.

Objective 8 - Cooperate with existing testing and validation approaches in Europe (and worldwide), taking into account other industries and domains for harmonisation and standardisation purposes.

Objective 9 - Improve and strengthen the expert network on CCAM safety assessment created in HEADSTART, to gather multi-stakeholder needs relevant for the project implementation and disseminate and promote adoption of the project results. This will be achieved through the development and establishment of different cooperation tools available to the participants.

Following a common approach will be crucial for present and future activities regarding the testing and validation of CCAM systems, allowing to obtain standardised results, to improve analysis and comparability, hence maximising the societal impact of the introduction of CCAM systems.

2.2 Project Partners

SUNRISE brings together world leading capability from across Europe.

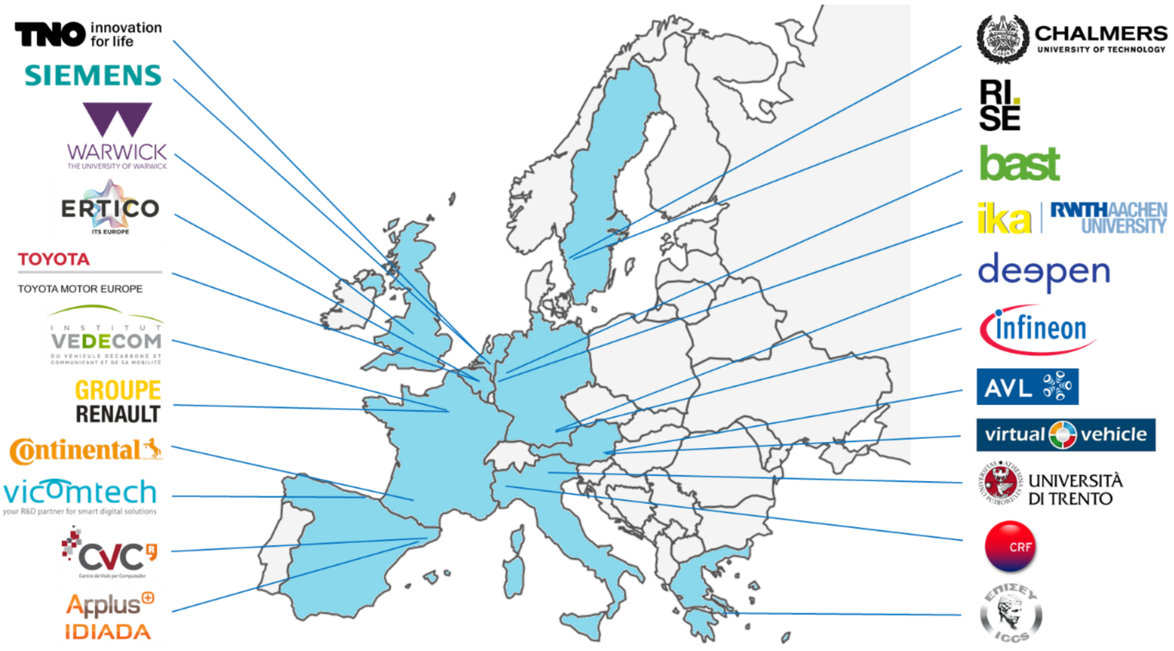

Figure 1: Project Partners

The full list of project partners is listed below:

- IDIADA Automotive Technology (IDIADA)
- AVL LIST GMBH (AVL)
- Bundesanstalt Fuer Strassenwesen (BASt)
- Continental Automotive France (CAF)
- Chalmers Tekniska Hoegskola (CHALMERS)
- Deepen AI (DEEPEN)

- European Road Transport Telematics Implementation Coordination Organisation (ERTICO)
- Institute of Communication and Computer Systems (ICCS)
- Infineon Technologies (IFAG)
- Rheinisch-westfaelische technische hochschule aachen (ika)
- Research Institutes of Sweden (RISE)
- Renault SAS (RSA)
- Siemens Industry Software Netherlands (SISW)
- Toyota Motor Europe (TME)
- Nederlandse organisatie voor toegepast natuurwetenschappelijk onderzoek (TNO)
- Institute VEDECOM (VED)
- Fundación Centro de Tecnologías de Interacción Visual y Comunicaciones Vicomtech (VICOM)
- Virtual Vehicle Research (ViF)
- Renault España (RESA)
- Universitá degli studi di Trento (UNITN)
- Centro Ricerche FIAT SCPA (CRF)
- Centre de Visió per Computador (CVC)

2.3 Work package structure

This highly ambitious project will be delivered through 9 co-ordinated work packages, which are summarised below:

- WP1 – Coordination and management – WP1 encompasses the coordination, management and governance of the project as well as other transversal activities that cannot be solved at WP level (e.g., exploitation, data management).
- WP2 – CCAM Safety Assurance Framework – will define how to derive safety evidence and arguments, with a scalable, economical, robust, interoperable and harmonised framework.
- WP3 – CCAM Verification & Validation (V&V) Methodology for Safety Assurance – aims to define and condense overall methodology to support the safety argumentation of CCAM systems.
- WP4 – CCAM V&V framework - WP4 develops a Harmonised V&V simulation framework and is connecting hybrid and real-world testing approaches for ODD-coverage based validation of CCAM systems.
- WP5 – Content harmonisation of scenario data framework – The WP5 activities will result in a future-proof set of commonly accepted and harmonised descriptions for the definition of the data framework content.
- WP6 – Data framework design and usage definition – will design and define the usage of the scenario data framework including the federation of the existing scenario databases and scalability to incorporate new SCDBs.
- WP7 – Use cases and framework demonstration instances development – Definition of the SUNRISE use cases (UCs) and associated CCAM validation requirements;

Development of the proposed framework Proof of Concepts (PoCs) for demonstration purposes realizing a selected set of SUNRISE UCs.

- WP8 – Cooperation with international vehicle safety bodies – will engage with the organizations that would later apply SUNRISE’s results, to make sure the results match their needs and can be applied in a flawless manner. This is done with direct contacts and in workshops. The vehicle safety bodies will be the EC especially DGs GROW & JRC, the UN ECE, Euro NCAP and the ISO.
- WP9 – Dissemination and international cooperation – will augment the project's impact and amplify its exploitation potential by laying out the overall dissemination and communication strategy, promoting SUNRISE’s activities and innovation to key stakeholder groups and facilitating international cooperation, including by establish an effective cooperation platform through the Expert Group

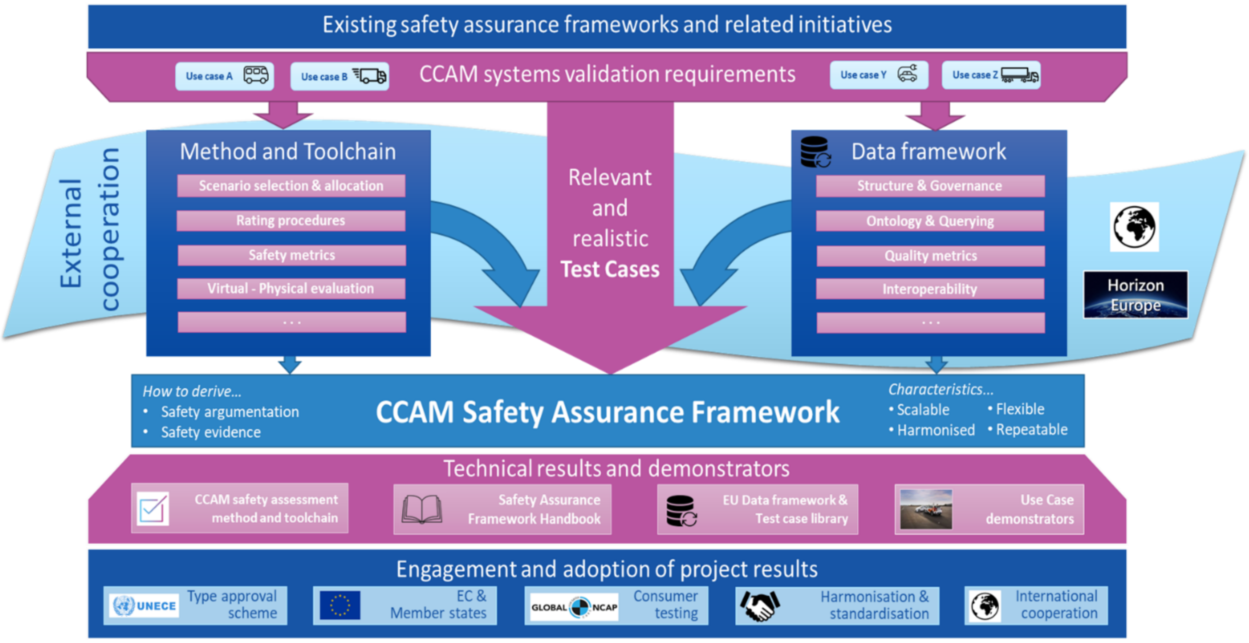
Existing safety assurance frameworks and related initiatives
Use case A
Use case B
CCAM systems validation requirements
Use case Y
Use case Z
External cooperation
Method and Toolchain
Scenario selection & allocation
Rating procedures
Safety metrics
Virtual - Physical evaluation
. . .
Relevant and realistic Test Cases
Data framework
Structure & Governance
Ontology & Querying
Quality metrics
Interoperability
. . .
Horizon Europe
How to derive...
Safety argumentation
Safety evidence
CCAM Safety Assurance Framework
Characteristics...
Scalable
Flexible
Harmonised
Repeatable
Technical results and demonstrators
CCAM safety assessment method and toolchain
Safety Assurance Framework Handbook
EU Data framework & Test case library
Use Case demonstrators
Engagement and adoption of project results
UNECE
Type approval scheme
EC & Member states
GLOBAL NCAP
Consumer testing
Harmonisation & standardisation
International cooperation

Figure 2: Work plan

3 SUNRISE DATA PRINCIPALS AND PROCESSES

Through the project proposal and the GA, the SUNRISE project and its partners commit to developing and implementing a DMP that:

'Defines procedures on how to handle personal data to guarantee project participants' fundamental rights and avoid misuse of the project results. It will specify how the data will be securely stored and whether it will be destroyed at the end of the project or archived for further use by the research community.' (SUNRISE proposal document)

And

The 'beneficiaries must promote the action and its results by providing targeted information to multiple audiences (including the media and the public)' (GA Article 17)

This section will identify the working principles of the project that will enable these commitments to be realised.

3.1 FAIR principals

In 2016, the 'FAIR Guiding Principles for scientific data management and stewardship' were published in Scientific Data. The authors intended to provide guidelines to improve the Findability, Accessibility, Interoperability, and Reuse of digital assets. The principles emphasise machine-actionability (i.e. the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data.

The FAIR principals, detailed guidance on each, and a wider set of supporting information can be found at https://www.go-fair.org/fair-principles/ . The principals are summarised below along with an explanation of SUNRISES commitment to align with these principals.

It is expected these principals will apply predominantly to data that is made publicly available. Nevertheless, the FAIR principles define good practice that should be followed where possible with all data across the project, at whatever level of availability.

3.1.1 Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the *FAIRification* process.

- (Meta)data are assigned a globally unique and persistent identifier
- Data are described with rich metadata
- Metadata clearly and explicitly include the identifier of the data they describe
- (Meta)data are registered or indexed in a searchable resource

3.1.2 Accessible

Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.

- (Meta)data are retrievable by their identifier using a standardised communications protocol
- The protocol is open, free, and universally implementable
- The protocol allows for an authentication and authorisation procedure, where necessary
- Metadata are accessible, even when the data are no longer available

3.1.3 Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

- (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- (Meta)data use vocabularies that follow FAIR principles
- (Meta)data include qualified references to other (meta)data

3.1.4 Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

- (Meta)data are richly described with a plurality of accurate and relevant attributes
- (Meta)data are released with a clear and accessible data usage license
- (Meta)data are associated with detailed provenance
- (Meta)data meet domain-relevant community standards

The SUNRISE project will follow the FAIR principals where possible, relative to commercial and practical restrictions.

3.2 Data management processes and procedure

The way in which data is handled and disseminated is of the upmost importance. The project brings together a wide range of partners with different internal processes and objectives, from commercial, to technology and research. It is essential that a common understanding of how data is handled, the conditions under which data is made public and the processes of agreeing such conditions is established and maintained. In doing so, a culture of trust can be established which in turn will enhance the collaboration of the partners, and ultimately the project outcomes.

Whilst the day-to-day interactions between partners and the appropriate employment of due diligence is what builds such trust, this section will provide guidance and recommendations on best practice.

3.2.1 Data Originator

Data Originator: The data originator is the organisation that originally created the data. Unless otherwise and explicitly stated it should be assumed the Data Originator owns the data and has the authority and responsibility to define who that data can be shared with, for what purpose and its classification.

The data originator should follow the best practice guidelines in this DMP and within their organisation when sharing data, including clear classification and labelling.

3.2.2 Expected data flow

- Figure 3: Data flows within and beyond the project' illustrates the expected data flow through the project.
- It is expected that most of the data will either flow into (from outside of the project) or be generated within the technical and co-operation work packages.

In all cases it is expected data will be handled with 'the same degree of care with regard to the Confidential Information disclosed within the scope of the Project as with its own confidential and/or proprietary information, but in no case less than reasonable care' CA, Section 10.5').

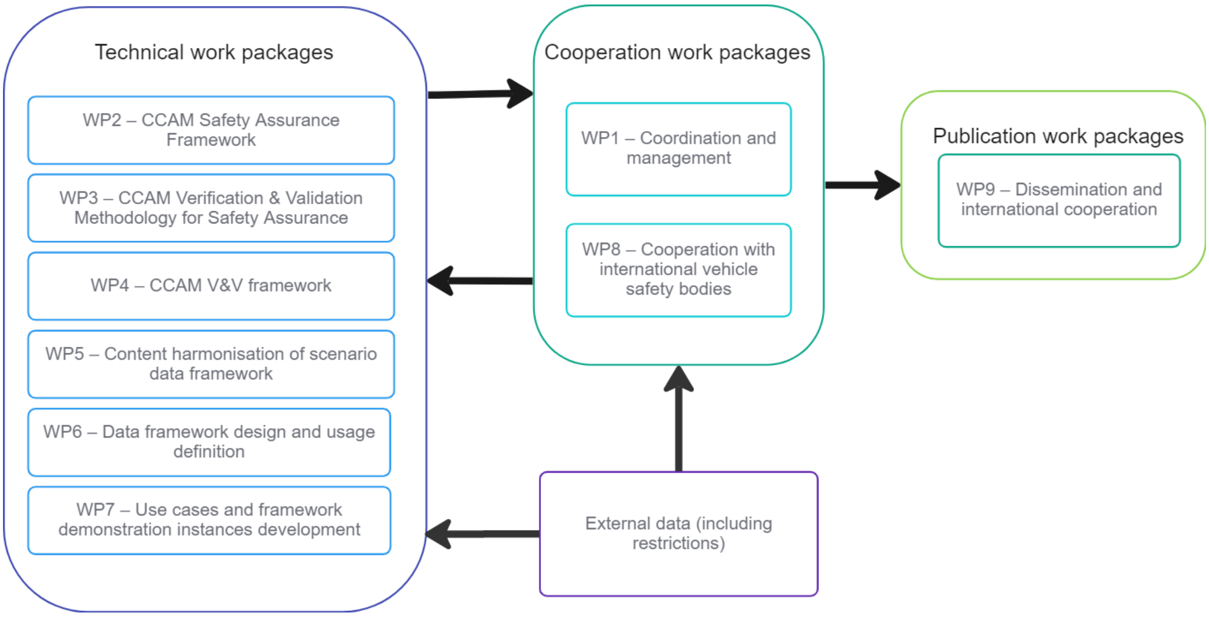
Technical work packages
WP2 – CCAM Safety Assurance Framework
WP3 – CCAM Verification & Validation Methodology for Safety Assurance
WP4 – CCAM V&V framework
WP5 – Content harmonisation of scenario data framework
WP6 – Data framework design and usage definition
WP7 – Use cases and framework demonstration instances development
Cooperation work packages
WP1 – Coordination and management
WP8 – Cooperation with international vehicle safety bodies
Publication work packages
WP9 – Dissemination and international cooperation
External data (including restrictions)

Figure 3: Data flows within and beyond the project

In order to transparently and effectively meet this obligation, the following sections provide recommendations and guidance on:

- Classifying data
- Labelling data
- Processes for public dissemination
- Key contact points

3.2.3 Classifying data

The simplest and most effective way of ensuring data is handled with care is a simple classification system. The SUNRISE project will have two levels of classification

- Sensitive or Confidential: Data classified as 'Sensitive' or 'Confidential' is limited under the conditions of the GA and CA. In this instance, the data originator should clearly label the data before sharing it. It is recommended this labelling is made bold and red. Furthermore, it is recommended the 'Labelling' guidelines below are followed to ensure the purpose for sharing the data is made clear to the recipient. It should not be assumed 'Sensitive/ Confidential' data can be shared with the whole consortia. It may be shared for the purpose of completing a work package or task. Any recipient of 'Sensitive/ Confidential' data should identify the data originator and gain their permission before sharing with other project partners or beyond the project. Data labelled as 'Sensitive/ Confidential' should not be shared with other project partners or outside the project without approval from the data originator. It should be noted that in certain circumstances data maybe shared outside of the project and remain sensitive/ confidential, for example to share with the Commission, or with Standardization Bodies. It is recommended that the data originator is always consulted, and that the labelling processes outlined below be followed so the intended recipients and usages are followed. **Note: This text is only intended as guidance, please refer to the relevant sections of the GA and CA.**
- Public : Data that is or can be put into the public domain. It is recommended that any data in whatever format and classification (or without classification) is verified as approved for public dissemination with the data originator before being made public.

For personal data and data that falls within GDPR, additional consideration maybe needed, please refer to the subsequent section of this DMP 'Personal data and GDPRPersonal data and GDPR'

3.2.4 Labelling

Within the project a multitude of datasets that cover several file formats, data types, scenarios, sizes, etc. will be produced. Each partner will contribute with either private or publicly available datasets. Whenever datasets are shared within the project, it is recommended that they are:

- Clearly classified (as above)
- Annotated according to the table below

Table 1: Data table to be completed for each dataset

Dataset name	**Name**
Dataset classification	Sensitive/ Confidential/ Public
Dataset reference	SUNRISE_WPX_TX.X_XX Each data set will have a reference that will be generated by the combination of the name of the project, the Work Package and Task in which it is generated and the reference number
Data set owner	Owner organisation(s) and contact person(s)
Organisations the data owner has provided authorisation for usage	The name(s) of organisations that the data has been shared with, and the intended usage of the data
Source	Dataset source (specify reuse if applicable)
Use case(s) & WP's	Related UC and/ or WP's in the SUNRISE project.
Data set description	Each data set will have a full data description explaining the data provenance, origin and usefulness. Reference may be made to existing data that could be reused.
Data format	All the format that defines data.
Data sharing	Explanation of the sharing policies related to the data set, including any associated terms and conditions (i.e., data generated outside of the project
Reuse of existing data	If applicable
Archiving and preservation	The preservation guarantee and the data storage during and after the project (databases, institutional repositories, public repositories…)

Annotating data with the above table will ensure full transparency with those the data has been shared with, and increase the likelihood data is not shared without the correct authority.

Where data is shared using standard Microsoft applications (or similar) i.e. Word, Excel, PowerPoint, the above table should be included in a clear primary position i.e. page 1 or 2 or a word document, the first or second slide on a PowerPoint or the first tab of an Excel. Where data is shared in niche formats i.e. cloud storage, simulation platform etc – it is recommended the content of the above table is communicated in a suitable fashion either within the system itself or clearly communicated before access is granted. In all situations it is recommended internal guidance also be requested and where appropriate additional safeguards such as password protection, 2-factor authentication and other appropriate measures relative to the sensitivity are employed.

3.2.5 Processes for public dissemination

The projects GA, Annex 5 states

The beneficiaries must disseminate their results as soon as feasible, in a publicly available format, subject to any restrictions due to the protection of intellectual property, security rules or legitimate interests.

Data will be shared among the project partners for a variety of reasons and purposes. For example, data maybe shared within or between work packages to enable certain technical tasks and deliverables to be achieved. Data maybe shared with project/ management level work packages to enable cross project communications, and data may be shared with the commission, standardization bodies or other stakeholders.

It is critical that a process is implemented whereby approval for a new use is always requested from the Data Originator. 'Figure 4: Data flow and processes' illustrates how new requests for data usage should always be sent to the Data Originator.

This is especially important when data moves from the project to the public domain. The GA and CA provide specific time scales and policies which must be followed before making data public.

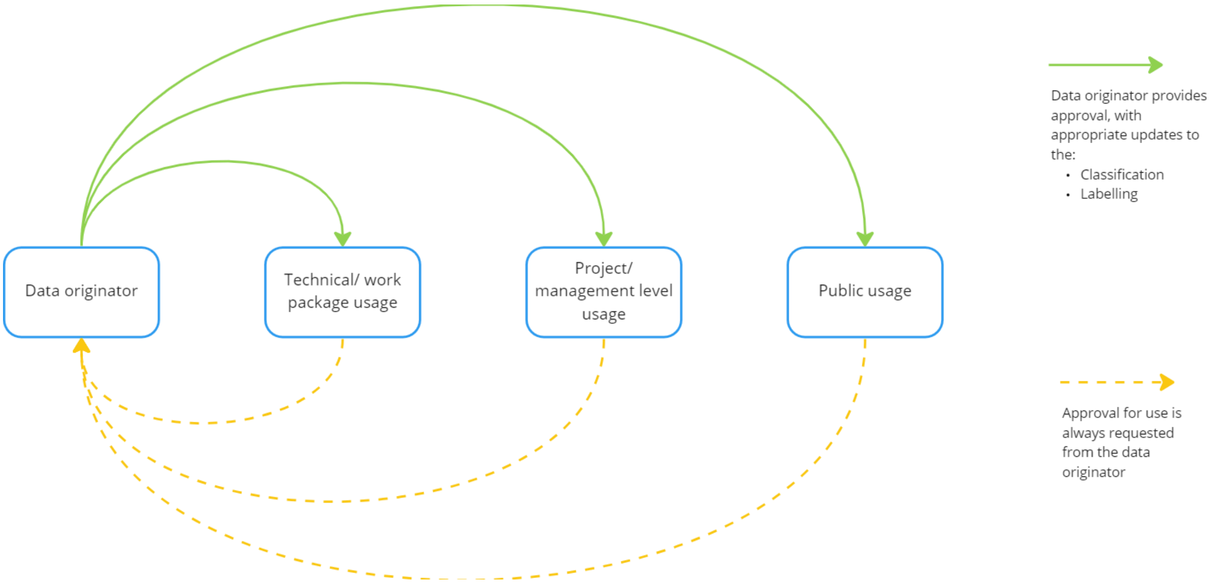
Data originator
Technical/ work package usage
Project/ management level usage
Public usage
Data originator provides approval, with appropriate updates to the:
• Classification
• Labelling
Approval for use is always requested from the data originator

Figure 4: Data flow and processes

3.3 Key contact points

When handling data within the project, it is always advised the CA and GA are consulted. In addition, key personnel within the project may be able to provide guidance. Key personnel include:

Liaison Manager (LM): The Liaison Manager is responsible for the exchange of relevant results and information with related projects or entities, with a special focus on cooperation with international vehicle safety bodies such as UNECE, Euro-NCAP, ASAM, ISO and SAE.

The SUNRISE project aims at providing approaches for safety validation, but to have an effect, the relevant stakeholders (as mentioned above) will need to be aware of these approaches, and more importantly, will need to be willing to adopt and apply them. Adoption of SUNRISE results by international vehicle safety bodies, is the main task of the Liaison Manager.

The role of Liaison Manager is fulfilled by Patrick Seiniger from BASt and elaborated mainly under WP8.

Dissemination Manager (DM): The main goal of the Dissemination Manager is to increase SUNRISE's impact by raising awareness and understanding of the project objectives and results among the stakeholder communities, and ensuring stakeholder engagement for fostering acceptance, endorsement and uptake of the project's results. Whereas the Liaison Manager mainly focusses on harmonisation with formal bodies, the Dissemination Manager interacts with the wider public and the CCAM community.

The role of Dissemination Manager is fulfilled by Gianmarco Donolato from ERT, and elaborated mainly under WP9.

Innovation and Exploitation Manager (I&EM): The Innovation and Exploitation Manager is responsible for the innovation management process. This includes the identification of innovations and their effective exploitation during and after the end of the project. The Innovation and Exploitation Manager guides the project to establish sound proposals based on specific technology evolution and CCAM validation maturation paths. The Innovation and Exploitation Manager will efficiently monitor market needs and technical evolutions, maximising the exploitation of the project's (final) results, while following the project Consortium Agreement.

The role of Innovation and Exploitation Manager is fulfilled by Thomas Schlömicher from AVL and elaborated mainly under WP1.

Data Protection Officer (DPO): The Data Protection Officer is the point of contact for data protection issues and coordination of the actions required to liaise between different partners and their respective Data Protection agencies and policies. The Data Protection Officer will ensure that data created, used and shared within the project and applies best practices and current regulation.

The role of Data Protection Officer is fulfilled by Siddartha Khastgir from UoW and elaborated mainly under WP1.

Work Package leaders: Work package leaders: Each work package has a designated leader. The organisational names can be found in the project proposal, and specific individuals on the BOX contact list. The WP1 leader (IDIADA) is responsible for overall SUNRISE project management.

4 DATA DESCRIPTIONS

4.1 Partner consultations

To better understand the data used and required in the project, members of the SUNRISE consortium were given a questionnaire about their data needs and their requirements for data handling. The questions asked were:

- What existing data do you expect to use within the project?
- What types of project data do you expect to generate?
- What internal data management policies do you have in place for handling data that is confidential or restricted?
- What internal data management policies do you have in place for making data available to the public?
- What are the key mechanisms and safeguards you'd like the projects DMP to cover?
- What standards for data management do you expect/ require the project to follow?

The responses to these questions served as input for different aspects of the DMP. Questions 1 and 2 were used to understand what data is expected to be covered in the project. Questions 3 and 4 were used to understand how the DMP would interact with internal data management within the organisations. Finally questions 5 and 6 were used to determine whether the DMP meets the expectations of the members of the consortium. The DMP covers all identified requirements and expectations.

4.2 Data themes

There is a variety of data that is expected to be used by the members of the consortium. This data may be either collected during the SUNRISE project or may have been generated prior. Most data are expected to fall under the following typologies, identified through the questionnaire:

- **Scenarios:** This is a specific type of data that is expressed through a language, such as ASAM OpenDRIVE/OpenSCENARIO and WMG SDL. A scenario contains data that can be used as input for virtual simulation tools, such as a description of the roads and environment, actors and their behaviour and traffic. This data can be generated from a multiple of sources, the most usual being derived from crash logs or using domain specific knowledge.
- **Simulation and test data:** This data is generated through various software and takes the form of simulation logs. Due to this, some of the data may not be interpreted directly without proprietary software, and not distributed as raw data.
- **Driver behaviour data**: This is data related to human drivers , and can be collected in a wide range of ways, varying from questionnaires to video data.*

- **Derived data:** This covers a variety of data, such as performance metrics, simulation results, and summaries.
- **Documentation:** This may be technical or not, and may contain information on how the other types of data were used to either help reproduce the results, or to show a rigorous process of obtaining the data.

As the nature and extent of these data themes will evolve during the project, more detailed descriptions might be provided in updated versions of the DMP.

4.3 Data types

The themes identified previously, manifest themselves through different data types. These data types are each linked with one or more working principles, which is presented in the following list:

- State of the art methodologies, such as literature, standards, analysis, white papers. This data will be contributed to the project by its range of expert partners, and in some instances will have been generated in previous EU projects such as HEADSTART and PEGASUS. These existing methodologies will be developed through the project along with the potential to develop novel approaches.
- Stakeholder engagement. This data is expected to be generated during the project across the requirements capture activities, and the dissemination activities. Data may be generated via questionnaires, surveys, interviews, the cooperation platform, or other communication tools. *
- Scenarios are a core element of the project and will be made up of both existing scenario data sets and new data sets. They will come in a range of formats such as SDL, OpenSCENARIO, OpenDRIVE.
- Simulation will be extensively undertaken across the project and its partners and will cover data such as scenario execution logs and simulation models.
- Physical test data will be generated from physical testing, this data will include video's, vehicle motion data*.
- Sensor data will be used in the project and will come data sensors such as source code, sensor models, and raw sensor data*.
- Driver behaviour data (such as facial features, driving behaviour) will be collected and analysed through the project from sources such as questionnaires and video.
- Observational data such as GPS, licence plates.
- Use Cases (UCs) describe a set of application specific CCAM systems V&V study cases. A SUNRISE UC is built upon a specific application based CCAM use case narrative, as the ERTRAC roadmap defines it (e.g., urban and sub-urban pilot), but its formal description is extended to also include the associated scenario space and validation profile assumed to be required for the UC.
- Derived or compiled data will be generated from the analysis of all the above data points.
- Interactive web-based handbook. This is a key public output for the project that will capture many of the public facing findings and outcomes.

- Publications. This covers all reports, publications, academic, industrial and press. This includes the planned interactive web-based handbook, academic publications and industrial reports.
- Financial and Administrative related information. This will be project specific data such as finances, project reporting, project member data, risks, and meeting minutes. *

A substantial component of the overall project is the creation of a data framework. This forms the deliverables in WP5 and WP6, with other work packages contributing. The specific data management policies developed within these WP's and for the framework itself fall outside of this DMP as they will require substantial and specific consideration. Nevertheless, the principles and guidelines set out within this DMP will act as an initial framework to inform the development of the data framework.

*Items that have the potential to include personal data, and which will comply with GDPR and other personal data handling obligations. Please refer to section 'Personal data and GDPR' for details

5 PERSONAL DATA AND GDPR

The SUNRISE project will comply with GDPR and/ or other relevant personal data laws and requirements. The CA (4.4) states:

In the event the Parties process personal data for the purpose of the project, each Party represents to comply with all applicable local, national or international laws and regulations, including the Regulation (EU) 2016/679 of the European Parliament and of the Council, General Data Protection Regulation (GDPR) or any other law or regulation applicable to the activities conducted by such Party (Applicable Laws).

In case an additional agreement related to the processing of personal data is deemed necessary according to the Applicable Laws, or desirable by Parties, the Parties shall enter into good faith negotiations to reach such an agreement. In particular, the Parties shall, where necessary, conclude a separate data transfer, data processing, data sharing and/or joint controller agreement before any data processing or data sharing takes place.

As part of compiling this DMP, all project participants were asked to complete a questionnaire. This has provided valuable insight to the data types the project forecasts to generate. As can been seen in the previous section 'Data Descriptions' several data types have the potential to include personal data.

Personally identifiable data could be very simple (i.e. a list of project contacts) or detailed and extensive (i.e. a systematic interview study with many stakeholders).

Whenever any project member intends on collecting, processing or using personally identifiable data of any kind GDPR and other legal compliance requirements must first be considered. Below an overview of GDPR considerations are provided, together with a set of further resources and support that can be accessed to help in final decision making.

5.1 Key definitions

The GDPR defines an array of legal terms at length. Below are some of the most important ones (https://gdpr.eu/what-is-gdpr/):

- **Personal data**: Personal data is any information that relates to an individual who can be directly or indirectly identified. Names and email addresses are obviously personal data. Location information, ethnicity, gender, biometric data, religious beliefs, web cookies, and political opinions can also be personal data. Pseudonymous data can also fall under the definition if it's relatively easy to ID someone from it.
- **Data processing**: Any action performed on data, whether automated or manual. The examples cited in the text include collecting, recording, organizing, structuring, storing, using, erasing… so basically anything.
- **Data subject**: The person whose data is processed. These are your customers or site visitors.

- **Data controller**: The person who decides why and how personal data will be processed. If you're an owner or employee in your organization who handles data, this is you.
- **Data processor**: A third party that processes personal data on behalf of a data controller. The GDPR has special rules for these individuals and organizations. They could include cloud servers or email service providers.

Please note that the 'Data Controller of any individual data set that requires GDPR consideration is the one responsible for ensure GDPR obligations are accurately covered.

5.2 Key principals

If you process data, you have to do so according to seven protection and accountability principles ((https://gdpr.eu/what-is-gdpr/):

- Lawfulness, fairness and transparency — Processing must be lawful, fair, and transparent to the data subject. One of the following 6 must apply (Art. 6 GDPR - Lawfulness of processing - GDPR.eu) :
    - Consent provided
    - Performance of a contract
    - Compliance with a legal obligation
    - Protection of the vital interests of the data subject
    - Performance of a task carried out in the public interest or in the exercise of official authority vested in the controller
    - Legitimate interests pursued by the controller or by a third party
- Purpose limitation — You must process data for the legitimate purposes specified explicitly to the data subject when you collected it.
- Data minimization — You should collect and process only as much data as absolutely necessary for the purposes specified.
- Accuracy — You must keep personal data accurate and up to date.
- Storage limitation — You may only store personally identifying data for as long as necessary for the specified purpose.
- Integrity and confidentiality — Processing must be done in such a way as to ensure appropriate security, integrity, and confidentiality (e.g. by using encryption). Please see section 7 Data storage and security of the DMP for more information
- Accountability — The data controller is responsible for being able to demonstrate GDPR compliance with all of these principles

Please note that a key responsibility of the 'Data Controller' is to document the GDPR decision making processes relating to their data set.

5.3 Accessing further guidance

If you are, or think you maybe, dealing with personally identifiable data you will be classed as a 'data controller'. In order to ensure GDPR compliance the following steps are recommended:

1. Consult with your organisational GDPR expert and other available organisational material

2. Consult relevant guidance material, such as GDPR.eu (and ico.org.uk in the UK)
3. GDPR.eu provides an accessible check list (GDPR compliance checklist - GDPR.eu) that will help establish the correct processes and procedures in managing personal data
4. Consult with the SUNRISE project consortium; with the primary contact being the Data Protection Officer and the Project lead (please see section 3 SUNRISE data Principals and processes)
5. Record and document your decision making

6 OPEN DATA

The SUNRISE consortium is committed to providing benefit to the European Community in terms of open and fair access to scientific knowledge, standardization, and economic impact.

Through its dissemination plan the project will aim to make findings and data known and available to specific stakeholders such at industry, policy makers, and standardization bodies. Furthermore, industry associations, stakeholder workshops and wider communications channels will be leveraged to enhance the project outcomes.

Within all these domains the data made available will adhere to certain standards and principles. This will ensure the data meets the FAIR standards of being findable, accessible, interoperable, and reusable (please see section 4.1 FAIR principals). This is not only critical for its use by the intended target stakeholder groups, but also so a wider set of stakeholders from academic, industry and policy have the potential of making use of the data. The Data Protection Officer can provide guidance and help in project partners meeting these requirements.

The guidelines the project will follow are detailed below (GA, Annex 5):

- Produce a data management plan (DMP) – this document
- Follow the FAIR principals – 'please see section 4.1 FAIR principals of this DMP
- All the relevant peer-reviewed publications relating to the project results will be made available to the Open Research Europe (ORE) (EC's open access publishing platform: https://open-research-europe.ec.europa.eu/).
- A machine-readable electronic copy of the published version or the final peer-reviewed manuscript accepted for publication, is deposited in a trusted repository for scientific publications (at the latest at the time of publication).
- Immediate open access is provided to the deposited publication via the repository, under the latest available version of the Creative Commons Attribution International Public Licence (CC BY) or a licence with equivalent rights; for monographs and other Long-text formats, the licence may exclude commercial uses and derivative works (e.g.CC BY-NC, CC BY-ND)
- Information is given via the repository about any research output, or any other tools and instruments needed to validate the conclusions of the scientific publication. Beneficiaries (or authors) must retain sufficient intellectual property rights to comply with the open access requirements.
- Metadata of deposited publications must be open under a Creative Common Public Domain. Dedication (CC 0) or equivalent, in line with the FAIR principles (in particular machine-actionable) and provide information at least about the following: publication (author(s), title, date of publication, publication venue); Horizon Europe or Euratom funding; grant project name, acronym and number; licensing terms; persistent identifiers for the publication, the authors involved in the action and, if possible, for their organisations and the grant. Where applicable, the metadata must include persistent

identifiers for any research output, or any other tools and instruments needed to validate the conclusions of the publication.

As soon as partners are aware data and findings are of public value they will begin, in accordance with the guidance in this plan and the terms of the CA and GA, exploring the possibility of making that data public. In some instances, the time required to making data available, could be subject to external influences i.e. internal Governance processes or Peer Review. Data will be made available '*as soon as feasible' (GA, Annex 5)*.

7 DATA STORAGE AND SECURITY

7.1 Best practice

The SUNRISE project is forecast to use a number of data depositories. A secure file sharing and working area for the project partners will be established for the day-to-day operations. However as bespoke software, data and systems are used and generated (i.e. simulation platforms, scenarios, test data) a wider set of depositaries are expected to be used. In all instances project partners will follow the following best practice in selecting, implementing and maintaining data storage facilities:

- **Access Control**: SUNRISE dataset repository must be capable of controlling the level of access that each user has depending on their role. There must be appropriate mechanisms to define and enforce such access control (e.g., firewalls, file systems permissions, secure log-in, password renewal) including physical control. This is provided by the IT department of the repository maintainer or by the dataset repository technology itself.
- **Authentication:** It must guarantee that the system being accessed is the intended one and that the user is who he or she claims to be. During the project, the partners will have access using a private password. Once the datasets become public, an e-mail-based authentication mechanism will grant access.
- **Non-Repudiation:** To ensure the capability to prevent users from denying that data files were accessed, altered, or deleted, auditing processes must be implemented. This is provided by the IT department of the repository maintainer or by the dataset repository technology itself.
- **Data Confidentiality:** Within the scope of the project, the protection of information from unauthorized access and disclosure must be preserved by restricting per-user access and encrypting the information during transmission and also during storage. After the defined retention period expires, information erasure or destruction must be ensured. This is provided by the IT department of the repository maintainer or by the dataset repository technology itself.
- **Communication Security:** Communication only flows through encrypted communication channels. This is provided by the IT department of the repository maintainer or by the dataset repository technology itself.
- **Data Integrity:** SUNRISE must protect data from unauthorized, uncontrolled, or accidental alteration during storage or transmission with the use of checksum values, hash functions and digital signatures. This is provided by the IT department of the repository maintainer or by the dataset repository technology itself.
- **Availability:** Back-up mechanisms are a desirable property, mainly to avoid Denial of Service (DoS) attacks. This is provided by the IT department of the dataset maintainer or by the dataset repository technology itself.

The SUNRISE project has implemented a shared repository for information and documents. This shared repository uses the BOX service, and it is protected by usernames and passwords so that only members that have been granted permission can access it.

7.2 Data legacy

Before the end of the SUNRISE project the dissemination and communications activities will establish a legacy plan for the projects outcomes. It is expected that this will include data sets that will be of value to the consortium and wider eco-system beyond the life of the project. For this long term preservation and curation of data, before the end of the project, an appropriate data repository (or repositories) will be identified and used. Established protocols for making such data sets available for public use such as Re3data.org, ZENODO and OpenAIRE with guidelines on how to select such repositories, will be followed.

8 CONCLUSION

The Consortium Agreement (CA) includes provision for data management as agreed between the project partners, and the Grant Agreement (GA) include provisions for data management obligations as agreed with the European Commission. However, good data management also requires best practice guidance and project level processes.

This DMP provides this guidance and best practice.

Execution of the DMP is the responsibility of all project members relative to their roles and responsibilities. The Project Coordinator, WP leaders and Task leaders, under the guidance and supervision of the DPO, will provide wide reaching project support to implement the DMP and provide additional support where required.

It should be noted that the DMP is in no way a substitute for the project's legal documents and do not replace their enforcement in any way – if in doubt the legally binding documentation takes precedence.

The DMP identifies **the FAIR principals** as a key guiding point to ensuring data is findable, accessible, interoperable and repeatable. The principals are summarised in the DMP along with an explanation of SUNRISE's commitment to align with these principals. The SUNRISE project will follow the FAIR principals where possible, relative to commercial and practical restrictions

The way in which data is handled and disseminated, is of the upmost importance. The project brings together a wide range of partners with different internal processes and objectives, from commercial, to technology and research. It is essential that a common understanding of how data is handled, the conditions under which data is made public and the processes of agreeing such conditions is established and maintained. In doing so, a culture of trust can be established which in turn will enhance the collaboration of the partners, and ultimately the project outcomes. Whilst the day-to-day interactions between partners and the appropriate employment of due diligence is what builds such trust, the DMP provides guidance and recommendations on best practice.

This includes guidance on:

Identifying **Data Originators**: The data originator is the organisation that originally created the data. Unless otherwise and explicitly stated it should be assumed the Data Originator owns the data and has the authority and responsibility to define who that data can be shared with, for what purpose and its classification.

Best practice processes for managing data including:

- Classifying data
- Labelling data
- Processes for public dissemination

- Key contact points

In the preparation of the DMP, all project partners were consulted via a questionnaire. The structure of this questionnaire together with its outcomes, in the form of a definition of data themes and data types are provided.

It should be noted that a substantial component of the overall project is the creation of a data framework. This forms the deliverables in WP5 and WP6, with other work packages contributing. The specific data management policies developed within these WP's and for the framework itself fall outside of this DMP, as they will require substantial and specific consideration. Nevertheless, the principles and guidelines set out within this DMP, will act as an initial framework to inform the development of the data framework.

If you are, or think you maybe, dealing with **personally identifiable data** you will be classed as a 'data controller' under GDPR. Whenever any project member intends on collecting, processing, or using personally identifiable data of any kind, GDPR and other legal compliance requirements must first be considered. This DMP provides:

- The key definitions (Personal data, Data processing, Data subject, Data controller, Data processor)
- The key principals that should be followed (Lawfulness, fairness and transparency, Purpose limitation, Data minimization, Accuracy, Storage limitation, Integrity and confidentiality, Accountability)
- Accessing further guidance

As soon as partners are aware data and findings are of public value, they will begin, in accordance with the guidance in this plan and the terms of the CA and GA, exploring the possibility of **making that data public**. The DMP provides best practice guidance and processes for making data available within the project and in order that that data be maintained and available beyond the time frame of the project.

Finally, the SUNRISE project is forecast to use a number of data depositories. In all instances, project partners will follow best practices in selecting, implementing and maintaining data storage facilities.

9 REFERENCES

1. Gdpr.eu Proton AG (2023) Jan 2023 https://gdpr.eu/

2. go-fair.org GOFARI (2022) Dec 2022 https://www.go-fair.org/fair-principles/

3. Information Commissioner's Office www.ico.org.uk (2023) Jan 2023 Guide to the UK General Data Protection Regulation (UK GDPR) | ICO

4. SUNRISE Consortium (2022). Consortium Agreement

5. SUNRISE Consortium (2022). Grant Agreement.