SAFETY4RAILS

DATA MANAGEMENT PLAN

Deliverable 9.5

Lead Author: UNEW

Contributors: MDM, Fraunhofer, CEIS, STAM, IC, RMIT, UIC, UREAD, EOS + external reviewer

Dissemination level: PU - Public

Security Assessment Control: passed (originally for the D1.6)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 883532.

D1.6 DATA CONTROL AND MANAGEMENT PLAN			
Deliverable nr.:	9.5		
Version:	1.0		
Delivery date:	31/03/2021		
Deliverable due date:	31/03/2021		
Dissemination level:	PU - Public		
Nature:	Report		
Main author(s)	Emmanuel Matsika (UNEW)		
Contributor(s) to main deliverable production	Antonio De Santiago Laporte Katharina Ross Stephen Crabbe Florence Ferrando Davide Ottonello Eros Cazzato Natalie Miller Nader Naderpajouh Virginie Papillault Atta Badii Elodie Reuge	MDM Fraunhofer Fraunhofer CEIS STAM IC FRAUNHOFER RMIT UIC UREAD EOS	*Data Controller* *Team* *WP1 and WP11 Leader* *WP2 Leader* *WP3 Leader* *WP4 Leader* *WP5 Leader* *WP7 Leader* *WP8 Leader* *WP9 Leader* *WP10 Leader*
Internal reviewer(s)	Andreas Georgakopoulos Antonio De Santiago Laporte[1] Atta Badii[1] Uli Siebold[1] Stephen Crabbe	WINGS MDM UREAD IC Fraunhofer	*Quality Manager* *Security Advisory Board* *Ethics Board* *Technical Manager* *Project Coordinator*
External reviewer(s)	Per-Erik Johannsson	(European CBRNE Centre, UMEA University)	

Document control			
Version	Date	Author(s)	Change(s)
0.1	3 Feb 2021	Emmanuel Matsika (UNEW)	1st Draft based on the deliverable D1.6 data Control and Management Plan as input.
0.2	5 Mar 2021	Emmanuel Matsika (UNEW)	2nd Draft
0.7	22 Mar 2021	Emmanuel Matsika (UNEW)	3rd Draft
0.8	24 Mar 2021	WP Leaders	Updates of Tables 2 and 3
0.9	25 Mar 2021	Emmanuel Matsika (UNEW) Stephen Crabbe (Fraunhofer EMI)	Revision to adapt D1.6 content from confidential to public reporting
0.9_FINAL	31 Mar 2021	Emmanuel Matsika (UNEW)	Update following comments from the Stephen Crabbe (Fraunhofer EMI) and the external reviewer.
1.0	31 Mar 2021	Stephen Crabbe (Fraunhofer)	Creation of V1.0 from 0.9_FINAL, plus minor updates.

DISCLAIMER AND COPYRIGHT

[1] For original D1.6 report.

ABOUT SAFETY4RAILS

SAFETY4RAILS is the acronym for the innovation project: **Data-based analysis for SAFETY and security protection FOR detection, prevention, mitigation and response in trans-modal metro and RAILway networkS**. Railways and Metros are safe, efficient, reliable and environmentally friendly mass carriers, and they are becoming even more important means of transportation given the need to address climate change. However, being such critical infrastructures turns metro and railway operators as well as related intermodal transport operators into attractive targets for cyber and/or physical attacks.**The SAFETY4RAILS project delivers methods and systems to increase the safety and recovery of track-based inter-city railway and intra-city metro transportation.** It addresses both cyber-only attacks (such as impact from WannaCry infections), physical-only attacks (such as the Madrid commuter trains bombing in 2014) and combined cyber-physical attacks, which are important emerging scenarios given increasing IoT infrastructure integration.

SAFETY4RAILS concentrates on rush hour rail transport scenarios where many passengers are using metros and railways to commute to work or attend mass events (e.g. large multi-venue sporting events such as the Olympics). When an incident occurs during heavy usage, metro and railway operators must consider many aspects to ensure passenger safety and security, e.g. carry out a threat analysis, maintain situation awareness, establish crisis communication and response, and they have to ensure that mitigation steps are taken and communicated to travellers and other users. **SAFETY4RAILS will improve the handling of such events through a holistic approach.** It will analyse the cyber-physical resilience of metro and railway systems and deliver mitigation strategies for an efficient response, and, in order to remain secure given everchanging novel emerging risks, it will facilitate continuous adaptation of the SAFETY4RAILS solution; this will be validated by two rail transport operators and the results will support the re-design of the final prototype.

TABLE OF CONTENT

List of tables

List of figures

No table of figures entries found.

Executive summary

This deliverable Data Management Plan (DMP) describes the methodology for data management that is being employed in the framework of the SAFETY4RAILS project. The methodology described aims to safeguard the sound management of the data collected and generated during the project's activities across their entire lifecycle, while also making them FAIR (findable, accessible, interoperable and re-usable) where relevant. Moreover, the DMP identifies the anticipated activities required for making data FAIR, outlines the provisions pertaining to their security as well as addresses the ethical aspects revolving around their collection/generation. While the data management issues reside in this document only, ethical issues are further elaborated in D9.1, D9.2 and D9.3.

The DMP remains a living document in the framework of SAFETY4RAILS and will be updated as needed throughout the course of the project considering its latest developments and available results, as D9.6 and D9.7. Ad hoc updates may also be made when necessary, with a view to delivering an accurate, up-to-date and comprehensive DMP before the completion of the project. This deliverable is delivered in M6 of the project and is updated based on the latest information available up to the month of delivery.

The DMP constitutes 7 chapters:

Chapter 1 (Introduction) provides introductory information about DMP, the context in which this has been elaborated as well as about its objectives and structure.

Chapter 2 (Data Summary) presents a summary of the data sets to be collected and/or generated during the activities of SAFETY4RAILS including the purpose of data collection/generation as well as types and formats. Additionally, it outlines its origin, expected volume and the stakeholders that may find it useful.

Chapter 3 (FAIR Data) describes the methodology that is applied in the framework of SAFETY4RAILS in order to safeguard the effective management of data across their entire lifecycle, making it also FAIR.

Chapter 4 (Allocation of Resources) estimates the resources required for maintaining a FAIR data curation, while also identifying data management responsibilities.

Chapter 5 (Data Security) outlines the data security strategy applied within the context of SAFETY4RAILS along with the respective secure storage solutions employed.

Chapter 6 (Ethical Aspects) addresses ethical aspects as well as other relevant considerations pertaining to the data collected/generated during the implementation of the project.

Chapter 7 (Conclusion) outlines what has been deduced and the next steps foreseen in the framework of the project with respect to its data control and data management plan.

1. Introduction

1.1 Overview

This Data Management Plan (DMP) is a structured guideline that describes the comprehensive lifecycle of data, from conception to storage, analysis, preservation, distribution and re-use scenarios. It is a follow on from Deliverable D1.6 (Data Control and Management Plan).

The D1.6 was completed in December 2020. This D9.5 report is based largely on the same contents as the D1.6. It includes only small updates to reflect developments in the three months up to March 2021 and its public dissemination level. The D1.6 is a confidential report.

This deliverable is intended to help SAFETY4RAILS partners who will generate, store and use data to consider all relevant questions concerning all data generated during project activities. Such data may be content, metadata or software applications. In this version, software applications are largely still not described in detail. Partners will ensure that consideration is made of the long-term accessibility and subsequent reusability of the data. All this is done in line with the Guidelines on Data Management in Horizon 2020 and according to the EU General Data Protection Regulation (GDPR) where relevant.

In addition, formulating and following the DMP paves the way for long-term accessibility and subsequent reusability of the digital assets. The DMP is a living document in the framework of SAFETY4RAILS and will be updated as needed throughout the course of the project considering its latest developments and available results.

In more detail, this DMP provides a description of what (kind of) data is collected along the entire lifecycle of the project. Furthermore, it describes how the data is processed both *during* the project and *after* its completion. The data should be meaningful, accountable and reliable.

This description includes statements about the origin of data, contextual allegations or statements, information surrounding the data collection process, infrastructure used to store and manage data, as well as information regarding the publication, citation, long-term accessibility and, if necessary, deletion of data during or after the research lifecycle. If personal data is processed, reference is made to documents handling legal and ethical aspects, including statements on data protection, terms of use, copyright attribution and exploitation rights for further reuse, and licensing.

1.2 Structure of the deliverable

This document includes the following additional chapters:

- Chapter 2 (Data Summary) presents a summary of the data to be collected or/and generated during the activities of SAFETY4RAILS including the purpose of its collection/generation as well as its types and formats. Additionally, it outlines its origin, expected volume and the stakeholders that may find it useful. At this early moment the project does not know the full extent of data that will be generated.
- Chapter 3 (FAIR Data) describes the methodology that is applied in the framework of SAFETY4RAILS in order to safeguard the effective management of data across their entire lifecycle, making it FAIR.
- Chapter 4 (Allocation of Resources) estimates the resources required for making the project's data FAIR, while also identifying data management responsibilities.
- Chapter 5 (Data Security) outlines the data security strategy applied within the context of SAFETY4RAILS along with the respective secure storage solutions employed.
- Chapter 6 (Ethical Aspects) addresses ethical aspects as well as other relevant considerations pertaining to the data collected/generated during the implementation of the project.
- Chapter 7 (Conclusion) outlines what has been deduced on the next steps foreseen in the framework of the project with respect to its data control and data management plan.

2. Data summary

SAFETY4RAILS intends to collect and generate meaningful non-sensitive and sensitive data. The former does not fall into any special categories of personal data as those described within the General Data Protection Regulation (GDPR). This data may be quantitative, qualitative or a blend of those in nature and will be analysed from a range of methodological perspectives with a view to producing insights that will successfully feed SAFETY4RAILS' activities, enabling it to deliver evidence-based results and ultimately achieve the objectives of the project. Sensitive data will include that which is described under the GDPR, and additionally, security classified data provided by end-users or law enforcement entities working on or associated with this project. It is however not expected for the project to work with data with official "classified data" i.e. data requiring a Personal Security Clearance (PSC) certificate or Facility Security Clearance (FSC) certificate. The project's approach is also to avoid the use of such data. The project coordinator and Security Advisory Board will monitor and, if necessary, review this approach with recommendations to the Project General Assembly (PGA). With that in mind, the second chapter of the Data Management Plan (DMP) starts by explaining the purpose for which this data will be collected/ generated and how it relates to SAFETY4RAILS. It proceeds by describing the different types and formats of this data as well as its origin and expected volume, before concluding with an overview of potential stakeholders for whom it may prove useful for re-use.

2.1 Purpose of the data collection/generation and its relation to the objectives of the project

To successfully meet its objectives and ensure the production of evidence-based results, SAFETY4RAILS entails several activities during which data will be collected/ generated. The purpose for which this data is collected/ generated is interrelated with the objective of the activity during which it is produced. These activities along with their objectives in the framework of SAFETY4RAILS are as follows:

On the one hand, statistical data are needed from the involved end-users to get information on the current capacity, e.g. how many trains are on the track, how many passengers are transported, what is their average time on the track, at how many stations do the trains stop, etc. Already during the proposal phase, such data were collected from the involved end users, to get an indication of which topics this data will cover see Table 1.

Besides statistical data, the project relies on meta-data about data sources to adapt interfaces, develop adapters to data sources and make the data accessible to the involved tools. Meta-data should describe the typical outcome of data sources, e.g. sensors, in a way that interfaces and algorithms can be adapted. This meta-data should include data provision frequency, data types and meaning of data.

In contrast to the meta-data the involved tools will need to have the actual data from the data sources. We foresee to have in this context two kinds of durability of data. Firstly, persistent data that we can use during the development and research work in WP3-7. This is re-used during the project phase whenever it is necessary, in particular to enhance algorithms and tools. For example, these data will be used in WP4 to train AI-based systems to better monitor the current situation and to identify better anomalies or to better forecast unforeseen events. In WP 5 these data will be used to simulate a railway and metro network to predict cascading effects. Secondly: volatile data that will come into play especially in WP8 during simulation exercises in which the Safety4Rails Information System (S4RIS) will be evaluated with real and/or simulated data. This data will be used in SAFETY4RAILS to further develop the 18 tools as a core within the SAFETY4RAILS Information System (SRIS) to match to the requirements of the involved end users.

Data collected during the SAFETY4RAILS project, relates to fulfilling its objectives through the various tasks. Table 2 shows the data types to be collected per task.

TABLE 1: STATISTICAL NUMBERS OF THE 4 INVOLVED METRO AND RAILWAY OPERATORS IN SAFETY4RAILS [2]

TOPIC	METRO MADRID / SPAIN	METRO ANKARA / TURKEY	RFI - ROMA TERMINI / ITALY	SRI LANKA RAILWAY / SRI LANKA
Typical rush hours per day	7h30 – 9h30 14h – 16h and 16h – 18h local time	07:00 – 09:30 16:00 – 20:00 Local time		6h30 – 8h30 15h30 – 18h30 Local time
Nº of passengers at that time per train	250.000 passeng./hour in peak hours. 2,88 people/m^2 in the trains	~1200 – 1800 ppl/train	850 trains/day 480.000 passengers/day	Peak - 3000 passengers (Approximately) Off Peak – 750 passengers (Approximately)
Frequency of trains per metro / railway line	Around 4 min	4,5 minutes		5 min (Rush hours) 1 1/2 hours (Normal hours)
Average delay per train per line	Around 2 min	6,11 seconds		15 min
Nº of stations per line	302 stations, from 7 (Line 11) to 33 (Line 1)	2 lines – 11 stations 1 line – 12 station 1 line – 9 station		368 stations (9 lines)
Average distance between stations	999 m	1250 m	Metro Bus Station Light Railway Taxi station	4 km (1500 km /368 stations)
Nº of intermodal transportation hubs	9 huge ones 20 small ones	6		1 (Makumbura, Bus Station and Railway Station) 1 Proposed, Pettah, Colombo
Nº of connecting stations	50	54 (+3 under construction)		

TABLE 2: DATA TO BE COLLECTED FOR VARIOUS TASKS

WP	WPL	Tasks	Type/Formats of Data
1	FRAUNHOFER	T1.1	Data generation/collection: • For administrative and financial management and reporting (i.e. managing and reporting on the Grant Agreement such as electronic documents, Emails, databases and presentations). • Foreseen formats: .docx, .pdf, xlsx, .mpp, .txt, .html, .jpeg / .png etc .pptx
		T1.2	Data generation/collection: • For technical management (i.e. focussed on the consistency of the overall technical solution developed in the project such as electronic documents, Emails, databases and presentations). • Foreseen formats: .docx, .pdf, xlsx, .mpp, .txt, .html, .pptx
		T1.3	Data generation/collection: • For scientific and quality management i.e. focussed on the quality of execution of workplan such as electronic documents, Emails, databases and presentations. • Foreseen formats: .docx, .pdf, xlsx, .mpp, .txt, .html, .pptx
		T1.4	Data generation/collection:

[2] https://www.railway-technology.com/features/europes-busiest-railway-stations/

WP	WPL	Tasks	Type/Formats of Data
			• For the development and testing of the SAFETY4RAILS Information System (S4RIS) such as electronic documents, Emails, databases and presentations. • As far as possible data is collected from operators and where possible past studies. The data to be collected/generated is not fully defined. • The formats of the data are also not yet fully defined but are likely to include docx, .pdf, xlsx, .txt, .html, .pptx, plus the other formats identified in the further WP sections.
		T1.5	Data generation/collection: • For the management of the Advisory Board and end users such as electronic documents, Emails, databases and presentations. • For organising and collecting input (such as their opinions) with regards to e.g. requirements used as input to the development of the S4RIS, best practices, design and use cases, feedback on evaluation and validation results. The data to be collected/generated is not fully defined. • The formats of the data are also not yet fully defined but are likely to include: docx, .pdf, xlsx, .txt, .html, .jpeg /.png etc, .pptx.
2	CEIS	T2.1	• Data and information related to end-user needs & requirements, and current information on end-user management systems and processes regarding threats and crisis management. • Minutes of end-user online workshops (under Chatham House Rule), yielding a confidential deliverable. • Notes from consultations with end-users & external stakeholders: online meetings, phone calls, questionnaire(s). • Electronic documents (docx, PPT, Excel, PDF)
		T2.2	• Open-source data collected for literature review on past failures. • Qualitative feedback from end-users on said failures. • Current data modelling methods used in the 17 tools. • End-user consultations, feedback and validation (detailed above in T2.1 activities) • Electronic documents (docx, PPT, Excel, PDF)
		T2.3	• Existing specifications of the 17 S4R tools. • End-user consultations, feedback and validation (detailed above in T2.1 activities). • Electronic documents (docx, PPT, Excel, PDF)
		T2.4	• Existing standardisation & certification requirements of the 17 S4R tools • Current standardisation & certification requirements from end-users • End-users consultations, feedback and validation (detailed above in T2.1 activities) • Electronic documents (docx, PPT, Excel, PDF)

WP	WPL	Tasks	Type/Formats of Data
		T2.5	• Data related to specific requirements from end-users acting in multi-modal environments (*foreseen)* • End-user consultations, feedback and validation (partly covered by T2.1 activities) • Electronic documents (docx, PPT, Excel, PDF)
3	STAM	T3.1	In collaboration with WP5: • Data related to Railways infrastructure and network, especially IT and OT components features and connections between them. • Data from literature and stakeholders experience about potential cyber and cyber-physical threats and attacks against Railways. • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
		T3.2	• Data related from sensors and security measures present in the Railway infrastructure, for instance sensors types, their functionalities, their ability to detect some types of phenomena or parameters of the system. • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
		T3.3	• Data obtained from the previous tasks T3.1 and T3.2 and from WP2, as well as data concerning the SECURAIL tool logic requirements. • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
		T3.4	• Data models coming from T3.3 • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
		T3.5	• Data obtained from the WP3 studies (incident levels, emergency and crisis scenarios, incident response, etc). • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
4	IC	T4.1	• Meta-Data about sensors, sensor data of the railway and IT infrastructure (mostly time series data) • At least the foreseen formats: docx, xlsx, .txt, .html, csv, json, xml.
		T4.2	• IoT, SCADA and related system installation details from railway operators as well as related data on vulnerabilities and protection from vulnerabilities, publicly available data from OSINT sources • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
		T4.3	• Meta-Data about sensors, sensor data • At least the foreseen formats: docx, xlsx, .txt, .html, csv, json, xml.
		T4.4	• Railway infrastructure data, e.g. grid topology, interconnection of infrastructure, interdependencies between different networks (railway network, electric power distribution network, and other networks). • At least the foreseen formats: docx, xlsx, .txt, .html, csv.

WP	WPL	Tasks	Type/Formats of Data
		T4.5	• All data from T4.1-T4.4, and recorded video data streams. • At least the foreseen formats: docx, xlsx, .txt, .html, csv, json, xml, HDS, HLS, CMAF HLS, Smooth Streaming, MPEG-DASH, RTMP, RTSP/RTP, SRT, WebRTC.
5	FRAUNHOFER	T5.1	• In collaboration with WP2: Data from existing literature regarding risks and vulnerabilities (risk description, occurrence, likelihood, potential targets, consequences) • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
		T5.2	• All data should come from WP2: data regarding the infrastructure that will be simulated, such as the infrastructure's 3D model, assets/components of the infrastructure and their corresponding characteristics (operation model, operation time), arrivals/departures schedules according to simulated scenario, estimated number of passengers arriving/leaving the infrastructure and typical behaviour description, estimated number of personnel and corresponding role description. • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
		T5.3	• All data should come from WP2: System components and attributes (repair time, connections, type and purpose, etc.), disruptive event types and effects, system functions. • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
		T5.4	• Data from WP2: Mitigation measures including best practices, existing and novel measures and their effects. Including type of measure, expected outcome, responsibility for execution and decision. • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
		T5.5	• Data from WP2 and WP4 and other WP5 tasks. • Network parameters (IP address, network name, time stamp, device name, etc.) together with the threats information types. • At least the foreseen formats: docx, xlsx, .txt, .html, csv
6	UNEW	T6.1	• To be determined as the project progresses. However, this will be a collection of inputs from T1.4, WP2, WP3, WP4, WP4, WP5, WP7 and WP9 and follow the operational interoperability guidelines. • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
		T6.2	• To be determined as the project progresses. However, this will be a collection of inputs from T1.4, WP2, WP3, WP4, WP4, WP5, WP7 and WP9 and follow the Technical interoperability guidelines. • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
		T6.3	• To be determined as the project progresses. However, this will be a collection of inputs from T1.4, WP2, WP3, WP4, WP4, WP5, WP7 and WP9 and integrated for S4RIS platform. • At least the foreseen formats: docx, xlsx, .txt, .html, csv.

WP	WPL	Tasks	Type/Formats of Data
		T6.4	• To be determined as the project progresses. However, this will be a collection of inputs from T1.4, WP2, WP3, WP4, WP4, WP5, WP7 and WP9 • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
7	RMIT	T7.1	Data collection for: • Developing the asset management model: Asset inventory, description of each element and sub-elements of the infrastructure system. • Developing the normal degradation model: Maintenance records of at least two consecutive inspections, maintenance and repair records (preferably more to further validate the model), including thresholds to take actions associated with maintenance, repair, rehabilitation and retrofits. • Developing the investment and budgetary model: Investment plans, Budget plans with discretization into maintenance, rehabilitation, repair, prioritisation and response mitigation, Asset management policies, Decision making process for allocation of budget for maintenance, repair and rehabilitation. • Foreseen formats: docx, xlsx, .txt, .html, csv
		T7.2	• Data collection for creation of an ontology of the system: asset inventory, personnel, devices, systems and facilities. • The formats of the data are also not yet fully defined but are likely to include docx, .pdf, xlsx, .txt, .html, .pptx
		T7.3	• Data generation/collection for establishing the profile of threats and budgetary implications such as budget plans, threat scenarios. The data to be collected/generated is not fully defined. • The formats of the data are also not yet fully defined but are likely to include docx, .pdf, xlsx, .txt, .html, .jpeg /.png etc, .pptx.
		T7.4	• Data collection for budgetary scenarios development and simulation: Budget plans, threat scenarios. • Data collection for developing the fault tree analysis: Inventory, location of asset elements and their contribution to system failure, GIS location of assets. The formats of the data are also not yet fully defined but are likely to include: docx, .pdf, xlsx, .txt, .html, .pptx
		T7.5	• Data generation/collection for budget optimization: budget plans, investment plans, required resilience levels. The data to be collected/generated is not fully defined. The formats of the data are also not yet fully defined but are likely to include: docx, .pdf, xlsx, .txt, .html, .jpeg /.png etc, .pptx.

WP	WPL	Tasks	Type/Formats of Data
8	UIC	T8.1	• Scenarios descriptions, evaluation criteria, test descriptions, various roles and tasks of the partners • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
		T8.2	• Data needed for the tests, technologies to be tested • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
		T8.3	• Collection of the feedbacks and KPI's • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
		T8.4	• Lessons learnt • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
9	UREAD	T9.1	• Textual (Use-Case, Use-Context Specification Data) • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
		T9.2	• Textual (Prototypical Scenarios Data) • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
		T9.3	• Textual (Regulatory Framework and Standards Data) • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
		T9.4	• Textual (Planning Information) • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
10	EOS	T10.1	• Personal data of the targeted stakeholders (email addresses, first and last name, country, type of organisation, region, gender) • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
		T10.2	• Personal data of the targeted stakeholders (email addresses, first and last name, country, type of organisation, region, gender) • List of publications and posts • Statistical data from webpages and social media pages • Scientific publications • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
		T10.3	• Personal data of the citizens to engage with (email addresses, first and last name, country, type of organisation, region, gender) • List of publications and posts • Statistical data from webpages and social media pages • Scientific publications • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
		T10.4	• Personal data of the targeted stakeholders (email addresses, first and last name, country, type of organisation, region, gender) • At least the foreseen formats: docx, xlsx, .txt, .html, csv.

WP	WPL	Tasks	Type/Formats of Data
		T10.5	• Data used will depend on each partner, and other data is yet to be defined. • At least the foreseen formats: docx, xlsx, .txt, .html, csv.
11	FRAUNHOFER		Data generation/collection: for fulfilling ethical requirements such as electronic documents, Emails, databases and presentations i.e. the presentation of information to demonstrate and report that the project fulfilled the requirements regarding, in summary: • Research participants: procedures/criteria identification/recruitment, informed consent procedures, informed consent forms, DPO contact details, findings on policy. • Opinions/approvals: by ethics committees and/or competent authorities for research with humans. • Personal data: DPO for partners collecting personal data at public locations, explanation of "data minimisation" principle, technical and organisational matters to safeguard rights and freedoms, security measures to prevent unauthorised access, anonymisation/pseudonymisation techniques and processing stage, consent and measures for further processing of previously collected data. • No misuse: Risk assessment and details on measures to prevent misuse of research findings. There will be a strong link to WP9. Foreseen formats: .docx, .pdf, xlsx, .txt, .html, .pptx.

2.2 Types and formats of collected/generated data

During the SAFETY4RAILS project, different types of data will be collected and generated, which can be described in many ways depending on the source and physical format of the data (as also already indicated in Table 2). Examples include created electronic text documents, spreadsheets, questionnaires and transcripts, among others. Other types of data would be in a format in which different data types (qualitative, quantitative, etc.) are stored. SAFETY4RAILS will have available easily accessible formats, such as post scripts (e.g. pdf, xps, etc.), machine readable formats (xml, html, json, etc.), spreadsheets (e.g. xlsx, csv, etc.), text documents (e.g. docx, rtf, etc.), compressed formats (e.g. rar, zip, etc.) or any other format required by the objectives and methodology of the activity within the frame of which it is produced, especially those that are software development based.

As far as possible, interoperable data formats will be applied, such as open formats (csv, pdf, zip, etc.), and/or machine-readable formats (such as xml, json, rdf, html, etc.). This data will make it easy for interested stakeholders to re-use, where made available outside the consortium. The type and formats of the data collected/generated in SAFETY4RAILS will be into three categories, identified to date, namely,

- Data collected/generated by direct input methods.
- Data from testing and training with SAFETY4RAILS developments.
- Data collected/generated from dissemination, communication and stakeholder engagement activities.

2.2.1 Data collected/generated through direct input methods

Within SAFETY4RAILS, direct input methods encompass methodologies for collecting and generating data through interactions between consortium partners and external stakeholders, with the latter providing data to the former. The identification and selection of suitable data subjects will be based on purposeful sampling.

The data collection involving consortium partners and external stakeholders (including from Advisory Board) will be collected respecting confidentiality criteria for sensitive data and will be anonymised where relevant. Online consultations involving end-user output will respect the Chatham House Rule (no attribution). Planned activities include online workshops, online meetings, consultations thoughts, electronic communications, phone calls, questionnaires and interviews. SAFETY4RAILS collects quantitative and qualitative data from end-users and relevant practitioners and stakeholders to be used to guide the initial stages of activities, review, feedback and validate SAFETY4RAILS results. These activities are conducted primarily through WP1 (T1.5), WP2 and WP8.

Data will also be generated by consortium partners during the actual development and testing activities.

2.2.2 Data collected/generated by users of the SAFETY4RAILS Platform during Testing, Implementation and Training

SAFETY4RAILS will develop, as a core result, an application software platform (in WP6) integrating Risk Assessment (WP3), Monitoring (WP4), Simulation (WP5) and Policy and Investment (WP7). During testing, implementation and end-user training, data will be generated from the following sectors:

- Metro and Railway Sector
- Software development and ICT
- Academic/research sector
- Safety and Security
- Law enforcement

2.2.3 Data collected/generated from dissemination, communication and stakeholder engagement activities

2.2.3.1 Social media statistics (including Twitter, LinkedIn, Facebook, YouTube)

This data will be collected/ generated through a periodic monitoring of the project's social media statistics (including Twitter, LinkedIn, Facebook and YouTube as relevant) with a view to measuring and assessing the performance and results of the project's social media activity in terms of dissemination and communication. With that in mind, the data will be both qualitative as well as quantitative in nature addressing the metrics reached on each channel (e.g. followers, tweets impressions on twitter, friends, etc.). Additionally, this data will be followed by an analysis of the results stemming from it and possible ways to improve the results to reach the project's targets. All in all, the data will be stored in a Microsoft excel file (.xlsx) while at the same time the analysis of the results will be stored in a standard word document (.docx).

2.2.3.2 Data collected from project events (e.g. workshops, stakeholder engagement events, etc)

This data will be collected in two ways during the project, i.e.:

- Stakeholder engagement organised by SAFETY4RAILS (such as the final conference and other public events) consisting of the participants lists that will include participant location (e.g. city).
- The participation of SAFETY4RAILS consortium partners in third party relevant events to reach out and engage stakeholders, thus includes general information about the events attended and their outreach.

This data is collected to keep track of the results of stakeholder engagement activities and provide the opportunity to project partners to report on these activities. Moreover, this data will be updated every time a partner attends an event, or a partner organises an event. Finally, the data will be both quantitative and qualitative in nature and will be stored in a standard spreadsheet (.xlsx).

2.2.3.3 Newsletter subscriptions (e.g. contact details of subscribers)

A subscription form hosted in the project's web portal will aid the collection of this data in which any interested stakeholder can freely provide his/her contact details in a dedicated sign-up form to receive the most up-to-date news and outcomes of the project. With that in mind, this data will be collected so as interested stakeholders can be informed about the SAFETY4RAILS project and training activities. The data will be comprised of a list of stakeholders along with their personal information. It is expected to include the following information:

- Email address (required)
- First and last name (required)
- Country (requested)
- Type of organisation (requested)
- Region (requested)
- Gender (requested)

A copy of this contact list will be stored to the Newsletter email server which will be used for email campaigns and newsletters distribution. All personal information included in this contact list will be used and protected according to email server's Privacy Policy.

2.2.3.4 Data from dissemination and communication

This data will be collected through a periodic monitoring of the project's miscellaneous dissemination activities such as publications in relevant journals, posts in the blogs, etc. The data will consist of a list of publications and posts published by the consortium partners. The purpose of collecting this data is to assess the outreach and efficiency of the communication and dissemination activities during the implementation of the project which will also be part of the periodic reporting to the Research Executive Agency (REA). For this purpose, a template is foreseen to be shared with all partners to recommend activities to be performed and log the activities they performed. Finally, all the data will be integrated in a single excel file (.xlsx).

2.3 Origin of data and Re-use of pre-existing data

In SAFETY4RAILS, new data will be collected/generated by consortium partners as well as external stakeholders participating in the activities of the project and/or during the end-user training activities as indicated above. In addition, external groups of stakeholders from which new data will originate include:

- Knowledge, technology and innovation solution providers (e.g. within academic institutions and their technology/knowledge transfer offices, non-university public research organisations, research and technology organisations, high-tech SMEs and large enterprises, etc.).
- Policy designers and implementers at regional, national and EU level (e.g. in regional/national/EU authorities, development agencies, etc.).
- Past EU Projects (such as FAIR Stations, RAMPART, SECUREMETRO, etc)

Any specific pre-existing data may be utilised within the context of the project as well. Data models such as CAD models, existing ontologies may be provided by project partners for the development of e.g. the S4RIS platform, (WP3, WP4, WP5 and WP6) and training activities (WP2 and WP8). In fact, SAFETY4RAILS partners are bringing together 18 tools most of which are at TRL 4 to 6. Those with pre-existing datasets will enhance the already populated environment. Other pre-exiting data is expected to come from railway infrastructure managers and train operators.

2.4 The expected size of the data to be managed

Table 3 presents the different activities planned to be implemented during the project in which data is collected/ generated, the types and formats of the data as well as the expected size of the data. Refer to Table 2 for the type of data and its format.

TABLE 3 : EXPECTED SIZE OF DATA

WP	Activity	Expected Size of Data
1	Financial and project management activities	• Estimate for all WP1 date collected/generated project internally for one user (e.g. Project Coordinator (PC) is > 30 GB. (After 1,5 months, PC has >1,5 GB project data. • Size of date from Advisers and end-users is still to be estimated.
2	Requirements, specifications and SAFETY4RAILS architecture design.	To be determined*
3	Development and Implementation of a multi-lingual Risk Assessment tool capable of dealing with both cyber and cyber-physical threats	• At least 100 types of entities modelling the Railway infrastructure and network • At least 50 types of relations among entities
4	To set-up AI-based algorithms to detect and forecast anomalies or events.	Estimated in the realm of less than a few GB's for time series data, covering timespans between a few weeks up to half a year. Recorded video data streams are much more storage demanding (compared to time series data) and depend highly on the resolution as well the file format. Video data streams could cover timespans from minutes to weeks.
5	To set-up AI-based algorithms to detect anomalies and forecast To catalogue vulnerable components within the railway system. To catalogue risks and vulnerabilities Catalogue of mitigation measures.	• Data from sensors covering timespans between a few weeks up to half a year. • Catalogue with a number of components and vulnerabilities. • Catalogue with a number of risks and vulnerabilities • Catalogue of best practice mitigation measures as well as novel ones
6	Integration and evaluation of software components from WP3, WP4, WP5 and WP7. And data from T1.4, WP2 & WP9	• The data collected/generated by the different tools in WP3, WP4, WP5 and WP7 will be managed centrally in WP6, but hosted by individual tool providers.
7	To set-up Central Asset Management System (CAMS) for financial budgetary elements considerations with resilience strategies.	Estimated for all the WP7 including the database of condition ratings, the investment model of the railway infrastructure, and database of transition matrices. ~ 30 GB.
8	Data collection to test the solutions KPI's for the evaluation of the solutions	To be determined*
9	Establish Ethical Compliance Assurance Framework Establish Crisis Communication Framework Guidelines	Small

WP	Activity	Expected Size of Data
	Legal Framework for Certification and Standardisation Data management plan	
10	Statistical data from website page and social media	To be determined*
11	Ethics	> 100MB

* - As at M6, some partners are still not able to determine some of the inputs. Updates to these will be made as the document develops into D9.6 and finally D9.7.

2.5 Data Utility - Beneficiaries

Data generated by SAFETY4RAILS should be of interest to a wide range of potential stakeholders. This is mainly because it covers metro and rail security and safety, including intermodality also within the Smart City context. Potential data beneficiaries include policy makers (EC, and national governments), law enforcement entities, standardisation agencies, rail infrastructure managers, train operators, academic/research institutions, implementers & funders, and of course the SAFETY4RAILS partners themselves. Stakeholders that may find the data to be collected/generated by the project along with the benefits that could arise for them by utilising this data, are outlined in Table 4.

TABLE 4 : DATA UTILITY

Stakeholder Group	Data Utility
Policy makers (EC, and national governments)	Treatment of cyber-physical threats may require development of new or review of existing policies and regulations on cyber and physical threats to transportation systems, particularly rail systems. Data from the project may help to develop new and or updated policies and regulations.
Law enforcement and first responder entities	The accuracy and effectiveness of the response of law enforcement and first responder entities relies on the accuracy of the data and information. The SAFETY4RAILS platform is foreseen to be a complex, but real-time dynamic system capable of providing timely information valuable also to these entities. Data in the project may help law enforcement and first responder entities to review the data they collect, analyse and communicate.
Rail infrastructure managers and train operators (users)	Security threats to the rail system target rail infrastructure and rolling stock. The data collected and generated by SAFETY4RAILS, should help them in future preparations to prevent and/or respond to threats.
Standardisation agencies	Throughout its duration, SAFETY4RAILS is set on collecting and producing data on the development of cyber-physical security platform for rail systems. This is novel, and therefore may require updating of software and security standards. In addition, the data generated in an operational system could potentially be used in monitoring/documenting that relevant standards (e.g. Network and Information systems Directive) are adhered to by users.
Academic and research institutions	Cyber security and physical security have traditionally been treated separately. With cyber-physical treatment being relatively still emerging, it has a large potential of further research not only in the EU, but also worldwide. Recognising this, SAFETY4RAILS data could provide researchers in the multi-disciplinary and cross-cutting field of cyber-physical

	security with valuable insights into how a platform such as S4RIS is developed, integrated and evaluated. With data generated from practical applications in the SAFETY4RAILS software development and training activities, interested researchers may re-use the data as a basis to replicate similar studies within the same or different contexts. They may also design and launch new studies, generating comparable research findings to further advance the field of cyber-physical security, beyond rail transportation.
Implementers & funders	Collected data on the evaluation of S4RIS, as well as identified best practices, could provide experts with reliable input to analyse the potential opportunities, successes (and failures) generated under such innovation actions. This can in turn help them gain a better understanding of what could drive successful security software platforms, supporting them in facilitating knowledge flows to and from their respective nations/regions.
SAFETY4RAILS Partners	The data collected/generated during SAFETY4RAILS is intrinsically important for project partners to produce evidence-based results and ultimately achieve the objectives of the project. Indeed, this data will enable the co-design, development, fine-tuning and validation of the project's innovation activities; the data will be used to design, improve, evaluate, and validate S4RIS platform. At the same time, this data should hold meaningful utility for project partners beyond the end of the project as well, enabling them to build and capitalise upon interesting ideas and opportunities that should emerge regarding the exploitation of the project results.

3. FAIR data

3.1 Making data findable, including provisions for metadata

As stated in Section 2, SAFETY4RAILS intends to collect/generate both non-sensitive and sensitive data. The latter will be divided into two – that which is described under the GDPR, and additionally, security classified data provided by end-users or law enforcement entities working on or associated with this project. In applying FAIR principles, data classification will be applied to determine which one is publishable and to which audience. Section 5 further elaborates on classification in relation to security data. However, as mentioned already above, it is not expected for the project to work with data with official "classified data" i.e. data requiring a Personal Security Clearance (PSC) certificate or Facility Security Clearance (FSC) certificate.

3.1.1 Data discoverability and identification mechanisms

The SAFETY4RAILS DMP aims to safeguard the sound management of the data collected and generated during the project activities across their entire lifecycle, while also making them FAIR where relevant. The project places special emphasis on enhancing the discoverability of relevant data collected/generated during its activities. Subsequently, the project follows a metadata-driven approach to increase the searchability of the data, while also facilitating its understanding and reuse. Metadata is defined as "data about data" or "information about information". It is the glue which links information and data across the World Wide Web, and the tool that helps people to discover, manage, describe, preserve, and build relationships with and between digital resources.

Three distinct types of metadata have been identified to date and are presented below:

- Descriptive metadata used to identify and describe collections and related information resources. Descriptive metadata at the local level helps with searching and retrieving. In an online environment, descriptive metadata helps to discover resources. In most circumstance it includes information such as the title, author, date, description, identifier, etc.
- Administrative metadata is used to facilitate the management of information resources. It is helpful for both short-term and long-term management and processing of data. This is information that will not usually be relevant to the public but will be essential for staff to manage collections internally. Such metadata may be location information, acquisition in-formation, etc.
- Structural metadata enables navigation and presentation of electronic resources. Its documents how the components of an item are organized. Examples of structural metadata could be the way in which pages are ordered to form chapters of a book, a photograph that is included in a manuscript or a scrapbook or the JPEG and TIF files that were created from the original photograph negative, linked together.

Bearing that in mind, relevant data produced/used during SAFETY4RAILS will be discoverable with metadata suitable to its content and format. The project will employ metadata standards to produce rich and consistent metadata to support the long-term discovery; use and integrity of its data (see Subsection 3.1.5 for more details on the metadata standards adopted by SAFETY4RAILS).

In parallel, to further increase data discoverability, the data produced by SAFETY4RAILS and deemed open for sharing and reuse, will be deposited in suitable infrastructure that serve the purposes. Such an infrastructure can be an open data repository. By employing this data repository, the data produced during the implementation of the project can be located by means of a standard identification mechanism. Indeed, SAFETY4RAILS will be able to assign globally resolvable Persistent Identifiers (PIDs) on any data uploaded to the repository. An identifier is a unique identification code that is applied to a dataset, so that it can be unambiguously referenced.

Datasets not uploaded to a repository will be deposited in a searchable resource (i.e. the web portal of the project) and will utilise well-tailored identification mechanisms as well, in the form of standard naming conventions that will safeguard their consistency and make them easily locatable for project partners within the framework of the project. The following subsection provides further details in this respect.

3.1.2 Naming Conventions

Following a consistent set of naming conventions in the development of the project's data files can greatly enhance their searchability. Therefore, SAFETY4RAILS has created consistent data file names that provide clues to their content, status, and versioning, while also increasing their discoverability. In doing so, project partners as well as interested stakeholders are able to easily identify a file as well as classify and sort it.

According to the UK Data Archive (UK Data Service, 2020) a best practice in naming convention is to create brief yet meaningful names for data files that facilitate classification. The naming convention should avoid the utilisation of spaces, dots and special characters (such as & or!), whereas the use of underscores is endorsed, to separate elements in the data file name and make them understandable. At the same time, versioning should be a part of a naming convention to clearly identify the changes and edits in a file.

To facilitate the reference of the datasets that will be produced during its implementation, SAFETY4RAILS has employed a standard naming convention that integrates versioning and will consider the possibility of creating multiple datasets during an activity that entails data collection/generation. Indeed, SAFETY4RAILS' naming convention addresses this last issue by employing a unique element that captures the number of datasets that are produced under the same activity.

In particular, the naming convention employed by the whole project is described below:

S4R _ [Name of Study] _ [Number of dataset] _ [Issue Date] _ [Version number]

- S4R: Short name of the project, SAFETY4RAILS.
- Name of Study: A short version of the name of the activity for which the dataset is created.
- Number of dataset: An indication of the number assigned to the dataset.
- Issue Date: The date on which the latest version of the dataset was modified (YYYYMMDD).
- Version number: The versioning number of a dataset.

Below are examples that demonstrate the naming structure applied in the context of SAFETY4RAILS. These examples are indicative and do not necessarily correspond to actual datasets.

- S4R_Project Management_Dataset2_20201109_v2 – The second dataset created for project management structure and related aspects. The last modification of this dataset, which in this case produced the second version of the dataset, was on the 9th of December 2020 (09/11/2020).
- S4R_Operational Interoperability_Dataset1_20201114_v1 – The first dataset generated within the framework of the WP6, Task 6.1 - Operational interoperability of S4RIS and logistics. This is the first version of the dataset that was last modified on the 14th of November 2020 (14/11/2020).

3.1.3 Search Key Word

The project's data has provided with search keywords with a view to optimising its re-use by interested stakeholders during its entire lifetime. Subsequently, the metadata standards employed by SAFETY4RAILS provides opportunities for tagging the data collected/generated and its content with keywords. The keywords are a subset of metadata and include words and phrases used to name data. For SAFETY4RAILS, keywords are used to add valuable information to the data collected/generated as well as to facilitate the description and interpretation of its content and value.

Along these lines, the project's strategy on keywords is underpinned by the following principles:

- The "who", the "what", the "when", the "where", and the "why" should be covered.
- Consistency among the different keyword tags needs to be ensured.
- Relevant, understandable and clear keywording ought to be sought.

In general, the keywords will comprise terms related to cyber security, physical security, rail transport, software development, integration & implementation, rail infrastructure manager and train operators. The keywords will accurately reflect the content of the datasets and avoid words used only once or twice within them.

3.1.4 Versioning

Versioning of information makes a revision of datasets uniquely identifiable and can be used to determine whether and how data changed over time and to define specifically which version the creators/editors are working with. Moreover, effective data versioning enables understanding if a newer version of a dataset is available and which are the changes between the different versions allowing for comparisons and preventing confusion. As such, a clear version number indicator is used in the naming convention of every data file produced during the SAFETY4RAILS to facilitate the identification of different versions. Once a version is superseded by a the latest one, it will be saved in an archive folder of the project coordinator's server (Fraunhofer EMI LiveLink Exchange) for reference.

3.1.5 Standards for metadata creation

SAFETY4RAILS employs standards for creating metadata for the data collected/generated by the project, with a view to describing it with rich metadata and thus improving their discoverability and searchability. As a result, effective searching, improved digital curation and easy sharing will be realized. In addition, the metadata standards applied enable the integration of metadata from a variety of sources into other technical systems.

Project data not available for re-use, will also be annotated with open and machine-readable metadata following the Dublin Core Metadata standard. The Dublin Core Metadata element set (covered by the international standard ISO 15836) is a standard which can be easily understood and implemented and as such, is one of the best-known metadata standards. It was originally developed as a core set of elements for describing the content of web pages and enabling their search and retrieval. See also section 3.3.

3.2 Making data openly accessible

3.2.1 Openly available and closed data

SAFETY4RAILS Project is part of the H2020 and aims to "make the data collected/generated openly available with as few restrictions as possible, while at the same time protecting sensitive data from inappropriate access" (Unige, 2020). In prioritising resources for making data openly available, there will be a focus on data primarily needed to validate the results presented in scientific publications (European Commission 2020) and/or deliverables. This being a security project, further consideration must be also made for data that may be classified, i.e. the project will voluntarily restrict release of such data (see Section 5 for more details).Thus, generally, the project will adopt the good practice encouraged by the Open Research Data Pilot (ORDP), namely that of making scientific data as open as possible and as closed as necessary. This calls for project partners to disseminate the project's data that have the potential to offer long-term value to external stakeholders and do not harm the confidentiality, commercial interests and/or privacy of the stakeholders (including the project partners) that contributed to the collection/generation of this data, with a view to maximising the beneficial impact of SAFETY4RAILS.

Only anonymised and aggregated data will be made open to ensure that data subjects (i.e. individual persons) cannot be identified in any reports, publications and/or datasets resulting from the project. The project partner, MDM, serving as the data controller will undertake all the necessary anonymisation procedures to ensure that the data subject is no longer identifiable. More details on data management responsibilities are provided in Section 4.2.

Therefore, it is important to keep in mind that during the process of data anonymisation, data identifiers need to be removed, generalised, aggregated or distorted. It is cardinal to differentiate between anonymisation and pseudonymisation, which falls under a distinct category in the GDPR (anonymisation makes the data subject unidentifiable, while pseudonymisation leaves room for the subject to be re-identified with additional information). To this effect, Table 5 provides a list of good practices for the anonymisation of quantitative and qualitative data derived from the tour guide on data management of the Consortium of European Social Science Data Archives (CESSDA, 2020).

Bearing this in mind, Table 6 presents the data to be collected/generated during the project that is foreseen to date to be made openly available. In case certain data cannot be shared (or need to be shared under restrictions), a justification for that choice will be provided.

TABLE 5 : DATA ANONYMISATION BEST PRACTICES

Type of Data	Good Practices
Quantitative data	• Removing or aggregate variables or reduce the precision or detailed textual meaning of a variable. • Aggregate or reduce the precision of a variable such as age or place of residence. Generally, report the lowest level of georeferencing that will not potentially breach respondent confidentiality. • Generalise the meaning of a detailed text variable by replacing potentially disclose free-text responses with more general text. *Restrict the upper or lower ranges of a continuous variable to hide outliers if the values for certain individuals are unusual or atypical within the wider group researched.*
Qualitative data	• Use pseudonyms or generic descriptors to edit identifying information, rather than blanking-out that information. • Plan anonymization at the time of transcription or initial write-up, (longitudinal studies may be an exception if relationships between waves of interviews need special attention for harmonised editing). • Use pseudonyms or replacements that are consistent within the research team and throughout the project. For example, using the same pseudonyms in publications and follow-up research. • Use 'search and replace' techniques carefully so that unintended changes are not made, and misspelt words are not missed. • Identify replacements in text clearly, for example with [brackets] or using XML tags such as <seg>word to be anonymised</seg>. • Create an anonymization log (also known as a de-anonymization key) of all replacements, aggregations or removals made and store such a log securely and separately from the anonymised data files.

All personal data collected/generated will be considered as closed data prior to their anonymisation and aggregation to safeguard the confidentiality of the data subjects. This is particularly important for data collected by end-users through sources such as Closed-Circuit Television (CCTV).

This data will be securely stored by the consortium partners that collected them to be preserved in their respective records only for as long as necessary for them to comply with their contractual obligations and no longer than 5 years, subject to review, from the project's completion. During this period, the personal and/or security data will be accessible only by authorised individuals of SAFETY4RAILS consortium partners as outlined in Section 5. After this period, the personal data will be deleted from the respective consortium partner's records.

TABLE 6 : DATA AVAILABILITY

WP	Data	Availability	Remarks
1	Electronic documents and presentations, including deliverables with a dissemination level: PU (Public)	Immediate for public deliverables. Storage will be up to 5yrs after the project completion	Dissemination level given by Description of Action (and Security Advisory Board confirmation) for deliverables. All information on an ad-hoc basis following project procedures to agree information can be released publicly.

WP	Data	Availability	Remarks
2	Electronic documents and presentations, including deliverables with a dissemination level: PU (Public)	Immediate for public deliverables. Storage will be up to 5yrs after the project completion	No names or email addresses will be made public as part of WP2
3	Deliverables Electronic documents and presentations, including deliverables with a dissemination level: PU (Public)	Immediate for public deliverables. Storage will be up to 5yrs after the project completion	D3.3, D3.4 and D3.6 are public deliverables, with D3.1, D3.2, and D3.5 are confidential.
4	Electronic documents and presentations, including deliverables with a dissemination level: PU (Public)	Immediate for public deliverables. Storage will be up to 5yrs after the project completion	N/A
5	Electronic documents and presentations, including deliverables with a dissemination level: PU (Public)	Public deliverables will be available immediately they are approved by the EC. Confidential deliverables will not be available. Email addressed will be used during the duration of the project.	D5.2, D5.4 and D5.6 are public deliverables while D5.1, D5.3 and D5.5 are confidential.
6	Electronic documents and presentations, including deliverables with a dissemination level: PU (Public)	Public deliverables will be available immediately they are approved by the EC. Confidential deliverables will not be available. Email addressed will be used during the duration of the project.	D6.1, D6.3 and D6.4 are public deliverables while D6.2 is confidential.
7	Electronic documents and presentations, including deliverables with a dissemination level: PU (Public)	No longer than 5 years from the project's completion	N/A
8	Data collection to test the solutions. KPI's for the evaluation of the solutions	No longer than 5 years from the project's completion	European Commission audits can occur within 5 years after the project end
9	Electronic documents and presentations, including deliverables with a dissemination level: PU (Public)	No longer than the duration of the project	Dissemination level given by Description of Action (and Security Advisory Board confirmation) for deliverables.
10	Electronic documents and presentations, including deliverables with a	Immediate for public deliverables. Storage will be up to 5yrs after the project completion	The WP10 will generate material with a primary purpose of informing the general public

WP	Data	Availability	Remarks
	dissemination level: PU (Public)		
11	N/A	Storage will be up to 5 years after the project completion	N/A

3.2.2 Data accessibility and availability

Public access to the open data will be made available through the project website (https://safety4rails.eu/) and/or an open access portal to be determined, which will automatically link to OpenAIRE. The data will be fully accessible. At the same time, closed data will be stored and shared amongst authorised members of the consortium through the web portals of the project. Table 7 presents where data will be made accessible in the context of SAFETY4RAILS.

TABLE 7 : DATA ACCESSIBILITY

WP	Data	Accessibility
1	All relevant confidential documents and presentations for the project such as deliverables, minutes of the meetings and teleconferences, Action Point lists etc.	Fraunhofer EMI LiveLink Exchange;
	Public deliverables and presentations	Project website and/or EC website.
	Data from end-users (through T1.5)	Fraunhofer EMI LiveLink Exchange (as relevant) and UIC Workspace platform
2	Data from end-users Deliverables (and draft/working versions), partners contributions, minutes of meetings and calls Publics deliverables	Project website Fraunhofer EMI LiveLink Exchange; Public website / EC website
3	Public reports on the development of the multi-lingual Risk Assessment tool and of the Crisis Management tool (D3.4 and D3.6)	Fraunhofer EMI LiveLink Exchange
4	Electronic documents and presentations, including deliverables with a dissemination level: PU (Public)	Fraunhofer EMI LiveLink Exchange. Other portals to be determined
5	Electronic documents and presentations, including deliverables with a dissemination level: CO	Fraunhofer EMI LiveLink Exchange
6	Data generated as a result of integrating the 18 software tools from WP3, WP4, WP5 and WP7. Additional data will come from T1.4, WP2, WP9.	Through the S4RIS platform. The platform architecture and therefore the graphical user interface are yet to be developed.
7	Public deliverables and presentations	Project website and/or EC website.
8	Data collection to test the solutions. KPI's for the evaluation of the solutions.	Partners who will test the S4RIS. Project deliverables through project website.

WP	Data	Accessibility
9	Metadata from Work package/Task Leaders	Restricted access to Consortium members only through SAFETY4RAILS project shared repository
10	Personal data of the stakeholders	Kept in a document protected with a password
11	Deliverables and presentations	Restricted access

3.2.3 Methods, software tools and documentation to access the data

SAFETY4RAILS emphasises the accessibility of the data collected/generated during the project; the goal is no specialised method, software tool and/or documentation should be needed to access the data. Stakeholders should be able to access the data by simply using their web browser (e.g. Mozilla, Google Chrome, Internet Explorer, Safari, etc.) through their computers (either desktop or laptop), smart phones and/or tablets. Closed data will be accessed only by authorised project partners through the respective member section of SAFETY4RAILS's web portals, hosted by the Project Coordinator (Fraunhofer), MDM and UIC. Again, no specialised method, software tool and/or documentation will be needed. Nevertheless, the member section of the web portal is accessible only with the insertion of a unique username and password combination.

3.2.4 Data, metadata, code and documentation repositories

SAFETY4RAILS's open data along with their linking metadata as well as any relevant code and documentation (if applicable) required to access this data, will be deposited to and securely stored by Fraunhofer EMI initially. Once a decision is made on the open repository to be used, SAFETY4RAILS data will be transferred and hosted by other suitable repositories without undue delay. Since all of project's openly available data will make use of PIDs (i.e. DOIs), the links to the data will not be affected. Meanwhile, SAFEY4RAILS's data that will not be openly available for sharing will be deposited, together with their accompanying metadata, code and documentation (if necessary), to the web portal of the project at Fraunhofer EMI. In addition, security-sensitive (or classified) data will be treated in line with the data control measures stipulated in Section 5.

3.2.5 Restrictions

When considering using open access portals for sharing the project's openly available data, SAFETY4RAILS will assess any potential restrictions that may apply (such as ethical, rules of personal data, intellectual property, commercial, privacy-related, security related, etc.). Specific restrictions depend on the data policy of the selected portal(s). Since SAFETY4RAILS has not yet determined the portal(s) to be use, this issue will be outlined in future DMP deliverables. Project partners will mainly use the open access level to disseminate the project's data amongst the interested stakeholders. However, there will be instances when embargo periods or restricted access may be used. Data that will not be available for re-use will be accessible only by authorised project partners and/or authorised personnel from the European Commission Services. In any case, SAFETY4RAILS will ensure open access to all peer-reviewed scientific publications that may be produced in the framework of the project, in accordance with the Grant Agreement.

This section has provided the methodology to ensure that the project data is as openly accessible as possible by any stakeholder that may find it beneficial for re-use. SAFETY4RAILS also focuses on providing metadata standards and appropriate metadata vocabularies to increase its data interoperability as elaborated in the following section.

3.3 Data Interoperability

Data interoperability refers to the ability of systems and services that create, exchange and use data to have clear, shared expectations for the contents, context and meaning of that data (Steele and Orrell, 2017). Based on this, SAFETY4RAILS has adopted in its data management methodology the use of metadata vocabularies,

standards and methods that will increase the interoperability of the data collected/generated through its activities (as described above).

For data that will not be publicly shared, the Dublin Core Metadata standard will be applied. This standard is a small "metadata element set" which accounts for issues that must be resolved in order to ensure that data meet traditional standards for quality and consistency, while at the same time, remaining broadly interoperable with other data sources. The elements of the standard provide a vocabulary of concepts with natural-language definitions that are instantly converted into open machine-readable formats (such as XML, HTML, etc.), enabling machine-processability. Table 8 shows the vocabulary of the Dublin Core Metadata (Sugimoto et al, 2002).

TABLE 8 : DUBLIN CORE METADATA STANDARD VOCABULARY (SUGIMOTO ET AL, 2002)

No.	Element	Element Definition
1	Title	A name given to the resource.
2	Creator	An entity primarily responsible for making the content of the resource.
3	Subject	The topic of the content of the resource.
4	Description	An account of the content of the resource.
5	Publisher	An entity responsible for making the resource available.
6	Contributor	An entity responsible for making contributions to the content of the resource.
7	Date	A date associated with an event in the life cycle of the resource
8	Type	The nature or genre of the content of the resource.
9	Format	The physical or digital manifestation of the resource.
10	Identifier	An unambiguous reference to the resource within a given context.
11	Source	A reference to a resource from which the present resource is derived.

The interoperability of openly available data will also be facilitated through an open access portal, with a metadata. This encloses HTML microdata that allows machine-readable data to be embedded in HTML documents in the form of nested groups of name-value pairs. The schema will provide a collection of shared vocabularies in microdata format that can be used to mark-up pages in ways that can be understood by the major search engines.

3.4 Increase data re-use (through clarifying licences)

3.4.1 License schemes to permit the widest use possible

In this section, licences are considered. These are instruments which permit a third-party to copy, distribute, display and/or modify the project's data only for the purposes that are set by the licence. Such permission is usually conditional. Although there are variations, three conditions are commonly found in licences which are the attribution, non-derivative, and non-commerciality. SAFETY4RAILS will publish its openly available data under the Creative Commons licencing scheme to foster their re-use and build an equitable and accessible environment for them. Different licensing schemes may be selected depending on the needs of the SAFETY4RAILS project, and the interests of the consortium generally, but also the rights of individuals for whom the data is about. Any updates to this section will be reflected in future DMPs accordingly in D9.6 and D9.7.

3.4.2 Availability for re-use

As previously stated, re-use of data is an important aspect in the SAFETY4RAILS methodology for making data FAIR. Sharing data to interested stakeholders will help in maximising the impact of the project on the EU citizens. It is expected that data will become available for re-use no later than 180 days after the end of its processing in the framework of the project (i.e. collection, anonymisation, aggregation, etc.). SAFETY4RAILS also recognises that there are partners who may seek to publish scientific results or apply for patents. In this case, these may request to postpone the public release of the data for up to two years. Nevertheless, it is also important to note that the period for which the data will remain available for re-use also depends on the restrictions of their repository. Table 9 shows the indicative expected time that SAFETY4RAILS data will be made openly accessible. This information will be updated in deliverables D9.6 and D9.7.

TABLE 9: EXPECTED TIME THAT DATA WILL BE MADE PUBLIC

WP	Name of Activity	Expected Date for making data Public	Remarks
1	Public deliverables	Immediately after they are released by the EC	As soon as the public deliverables are released by the EC, they will be published on the SAFETY4RAILS homepage.
	Public presentations	Once given	Added to website as relevant (not too many presentations with very similar content).
2	Public deliverables	Immediately after they are released by the EC	As soon as the public deliverables are released by the EC, they will be published on the SAFETY4RAILS homepage.
3	Public deliverables	Immediately after they are released by the EC	As soon as the public deliverables are released by the EC, they will be published on the SAFETY4RAILS homepage. Confidential deliverables will not be published.
4	Training and Test data for tools and algorithms	Within above mentioned 180 days after being rated as useful and confirmed by data owners	None.
5	Deliverables	Immediately after they are released by the EC	As soon as the public deliverables are released by the EC, they will be published on the SAFETY4RAILS homepage. Confidential deliverables will not be published.
6	S4RIS platform and related deliverables	D6.1, D6.2 and D6.3 – immediate. D6.2, remains confidential.	D6.1, D6.2 and D6.3 are public deliverables. Hence, they will be public as soon as the EC approves them. However, D6.2 will remain confidential.
7	Public deliverables	Immediately after they are released by the EC	As soon as the public deliverables are released by the EC, they will be published on the SAFETY4RAILS homepage.
8	Public deliverables	Immediately after they are released by the EC	As soon as the public deliverables are released by the EC, they will be published on the SAFETY4RAILS homepage.
9	Ethical deliverables	N/A	As soon as the public deliverables are released by the EC, they will be published on the SAFETY4RAILS homepage.

WP	Name of Activity	Expected Date for making data Public	Remarks
10	Statistics of activities on social media	At M12 and M24 for the first and second updates of the D10.1.	Data will be monitored at M12 and by M24.
11	Ethics	N/A	Ethics reports, which are confidential

3.4.3 Data quality assurance processes

Quality Assurance (QA) and Quality Control (QC) activities are intrinsic to SAFETY4RAILS's data management methodology through the Quality Assurance Plan (deliverable D1.5). Therefore, before any data is published, it is checked for quality. As such, SAFRETY4ARAILS safeguards the transparency, consistency, comparability, completeness and accuracy of the data.

QA is a planned system of review procedures conducted outside the framework of developing a dataset, by personnel not directly involved in the dataset development process (IPCC, 2006a). In SAFETY4RAILS it consists of peer-reviews of methods and/or data to assess the quality of the dataset and identify any need for improvement. It ensures that the dataset correctly incorporates the technical, scientific knowledge and data generated.

As part of the project activities, procedures will be woven in, designed to provide routine technical checks as they measure and control data consistency, integrity, correctness and completeness as well as to identify and address errors and omissions. Such checks will cover everything from data acquisition and handling, application of approved procedures and methods, and documentation. These include checking:

- The validity of the measurement methodology (where relevant);
- Confirmation of the correct implementation of the measurement/test methodology (where relevant);
- For transcription errors in data input;
- That scale measures are within the range of acceptable values;
- Whether proper naming conversions are used; and
- Any caveats to be included with the data

4. Allocation of resources

4.1 Anticipated costs for making data FAIR

The costs required for data collected/generated during SAFETY4RAILS activities FAIR are integrated within the budget of the project. The primary costs are personnel costs. These anticipated costs are needed to cover a set of specific data processing and data management activities. The data processing and data management activities are the following:

- Collection
- Checking Data Quality
- Documentation
- Storage
- Access and Security
- Preservation
- Availability and Reuse
- Overall Data Management

A description about each data processing or data management activity is given below. The “Collection”, “Documentation”, “Storage, access and security”, “Preservation” and “Availability and reuse” activities are part of the WP under which the respective data are processed so the required effort will be part of the respective WP. However, the overall data control and data management plan activity is part of T1.4 and WP9.

Collection covers all activities necessary for acquiring external datasets (if required), gathering/generating new data, transcribing (if applicable), formatting and organising this data as well as acquiring informed consent from data subjects. This activity accounts for most of the costs required for making data FAIR as the majority of SAFETY4RAILS data constitutes new data collected/generated over the course of the project. Data documentation costs address the effort required for describing data (e.g. marking data with variable and value labels, code descriptions, etc.) as well as creating well-defined metadata along with a meaningful description of the context and methodology of how data was collected/generated and processed (where necessary).

Costs for data storage include both the resources required for ensuring adequate storage space for the data as well as the effort necessary for conducting data back-ups, while data access and security costs encompass costs related to ensuring access to the data as well as for protecting it from unauthorised access or use or from disclosure. Given that the storage of most of SAFETY4RAILS’s data will not require the procurement of additional space (other than what is already available to project partners) as well as that no special measures or software are required to access and secure the data (other than that which is inherently built into the repositories of SAFETY4RAILS’s data), such costs are kept to a minimum. However, as elaborated in Section 5, for security sensitive (classified) data, designated security-coded external hard drives will be used located at the designated partner under lock and key. This will mitigate effects of security breaches on institutional servers. An example is when the Newcastle University system was hacked (ITPro, 2020).

Data preservation costs, on the other hand, are estimated relatively higher than data storage, access and security costs, as additional effort will be required in several cases in order to convert the collected/generated data from their original form (e.g. physical interview transcripts) to an open and/or machine readable format suitable for long-term preservation (e.g. to an .xlsx format.). Adequate effort for data availability and re-use costs is also foreseen to safeguard the appropriate digitisation and anonymisation of the data as well as cover any resources required for data sharing and cleaning. Along the same lines, appropriate effort is foreseen for overall data management as well, to cover the effort related to the operationalisation of data management.

Another anticipated cost can also be the fees that some publishers of academic and scientific journals charge to provide open access to articles or journals. Costs vary between different journals and publishers. Finally, costs for long-term preservation in SAFETY4RAILS are assumed to be negligible, although this will be re-assessed during the project. Any updates will be presented in D9.6 and D9.7.

4.2 Data management responsibilities

To effectively execute the SAFETY4RAILS DMP, specific data management roles have been assigned to various partners as follows (in reference also to the GA):

Data Controller (DC)
The DC (MDM) is responsible of the overall data management in the framework of the SAFETY4RAILS project, including the elaboration of the DMP and its update (when necessary and with support of all partners). Additionally, the DC is responsible of establishing and monitoring the procedures for the collection and usage of data within the project lifetime supported by all WP and Task Leaders and properly assisted by the Project Coordinator and the partner UNEW that is assisting the DC in the data control role. The DC will determine together with the generating/collecting partners the data that will be shared and will become publicly available in the appropriate platforms. It is also the responsibility of the DC to support the generating/collecting partner to ensure the required content and quality of the shared data.

Project Coordinator (PC)
The PC, Fraunhofer EMI, provides support to the DC in the execution of its responsibilities. Working with the End-Users Coordinator (EUC), UIC, the PC is responsible for checking that the project ethics requirements (in WP11) are met regarding e.g. the elaboration of suitable templates for the informed consent form and information sheet to be appropriately adjusted and utilised by project partners during the relevant activities of the project. This is in close cooperation with the WP9 and particularly the WP9 WPL and Ethical Manager UREAD. Finally, the PC together with the DC coordinates with Work Package and Task Leaders to determine when and where shared data becomes available.

Technical Manager (TM)
Besides the DC, the TM provides support to WPLs and WTLs to assure the availability of data in an appropriate amount and quality, so that they can fulfil their tasks in SAFETY4RAILS. The support particularly is for those partners who are providing tools. Where necessary, if primary data cannot be provided from the real world (e.g. real sensor data), the TM coordinates activities amongst the above-mentioned roles to make alternative data available. Such data might include secondary data from public sources that is compliant with the format that is required by the WPLs and WTLs. Furthermore, the usage of historic data for generation of representative data might be considered. However, that should be avoided whenever possible.

Work Package Leaders (WPLs)
The WPL is responsible for coordinating the implementation of the data processing activities performed under the WPs they are leading. They align with the PC and the respective Work package Task Leader on whether and how the data gathered/produced under the tasks that fall within the WP they are leading will be shared and/or re-used. This includes the definition of access procedures as well as potential embargo periods along with any necessary software and/or other tools which may be required for data sharing and re-use. Finally, the WPLs are responsible for assuring the quality of the data from the activities of the WP they are leading, including assessing their quality and indicating any need for improvement to the respective Work package Task Leaders.

Work package Task Leaders (WTL)
The WTL act as data controllers of the data collected/generated in their tasks. They determine the purposes and means of processing this data as well as safeguarding its appropriate and timely processing. In addition, they are responsible for properly adjusting the templates for the informed consent form and information sheet (where needed) to the needs and specificities of the activities carried out in the task they are leading. Finally, they undertake any necessary actions to prepare the data collected/generated through the tasks they are leading for sharing either within the consortium or openly (including the use of proper naming conventions, application of suitable anonymisation techniques, creation of appropriate metadata and documentation, etc.).

Data processors
Data processors are project partners that are tasked to collect, digitise, anonymise, store, destroy and/or otherwise process data for the specific purpose of the activity in which it has been collected/ generated within the framework of the project. They are responsible for appropriately collecting the necessary consent for processing data (where needed) as well as for ensuring that the informed consent form and information sheet used to this end are properly adjusted to the needs of the activity they are participating and any particularities applicable to their organisation. Additionally, they are also responsible for managing the consents they have retrieved with a view to demonstrating their compliance with the relevant applicable EU and national regulation.

Finally, they perform quality checks to assess and maintain the quality of the dataset(s) held within their records. MDM as DC will coordinate the data processors and help ensure that all partners implement the data management plan correctly.

Data repositories

Data repositories are tasked with the storage and long-term preservation of the project's data. This aspect will further be elaborated when SAFETY4RAILS determines which repository will be used for openly shared data. For day-to-day storage of data from various WPs, the PC's server, Livelink Exchange, will be used. Accordingly, the Web Portal of SAFETY4RAILS shall securely store and preserve the project's data available for sharing amongst authorised consortium members in the framework of the project.

5. Data security

As part of data control, SAFETY4RAILS will securely handle any collected/generated data throughout its entire lifecycle as it is essential to safeguard this data against accidental loss and/or unauthorised manipulation. Particularly, in case of personal data collection/generation it is crucial that this data can only be accessible by those authorised to do so. With that in mind, the project data security and Information assurance safeguards including backup and data recovery strategy aims at ensuring that no data breach or loss will occur during the course of the project and after the completion of SAFETY4RAILS, either from human error or hardware failure, as well as inhibit any unauthorised access.

All project partners responsible for processing data within their private servers will ensure that this data is protected, and any necessary data security controls have been implemented, to minimise the risk of information leak and destruction. This case refers to the data that is closed and therefore will not be shared and/or re-used within the framework of the project. In this case and to avoid data losses, the data will be backed up on a daily basis and the backed-up files will be stored securely in external hard disk drives so as to safeguard their preservation, while also enabling their recovery at any time. Additionally, integrity checks will be carried out at least once a month (or more often, if deemed necessary) ensuring that the stored data has not been changed or corrupted.

Access to closed or sensitive confidential or classified data is only permitted to authorised project partners. In the case that there is a personal or security data breach, within no later than 72 hours, project partners will notify national supervisory authorities (e.g. data protection authorities) as well as the data subject(s) that may have been affected by the breach. They will document any personal data breaches, including information such as the facts relevant to the breach, its effects and the remedial action(s) taken. As an additional feature, there is a recognition that identification and authentication access controls will play an important role in SAFETY4RAILS. This will help partners to protect the data collected/generated during the project. To this end, each project partner is responsible for and committed to ensuring the application of appropriate access controls to the data they are processing within the private servers of their organisation. At the same time, technical access controls are built into the web portal(s) of SAFETY4RAILS, setting out clear roles with access rights to the data stored there, so that only authorised personnel have access.

6. Ethical aspects

By virtue of its activities, SAFETY4RAILS involves collection, processing and generation of both meaningful non-sensitive and sensitive data. The former does not fall into any special category of personal data as those described within the General Data Protection Regulation (GDPR). Any personal data collected/generated in the framework of SAFETY4RAILS is processed according to the principles laid out by the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data which entered into force in May 2018 aiming to protect individuals' rights and freedoms in relation to the processing of their personal data, while also facilitating the free flow of such data within the European Union. In this project, data will be collected/generated only for specified, explicit and legitimate purposes relative to project objectives. Moreover, all project partners tasked with processing data during the project life will fully abide with their respective applicable national as well as EU regulations. The deliverable D9.1 SAFETY4RAILS Ethical Compliance Framework (ECF) provides further details on ethical aspects pertaining to the project including data (Badii et el, 2021).

7. Other issues

At this early stage of the project, there are no issues noted. As the project progresses, any issues that arise will be included in later versions of the DMP (D9.6 and D9.7).

8. Conclusion

8.1 Summary

This deliverable D9.5 has described the methodology for data management that is planned to be employed in the framework of the SAFETY4RAILS project. The methodology applied aims at safeguarding the sound management of the data collected and generated during the project's activities across their entire lifecycle, while also making them FAIR where relevant. Moreover, the DMP identifies the anticipated activities required for making data FAIR, outlines the provisions pertaining to their security as well as addresses the ethical aspects revolving around their collection/generation. The project places special emphasis on enhancing the discoverability of relevant data. Subsequently, it follows a metadata-driven approach to increase the searchability of the data, while also facilitating its understanding and reuse.

Both qualitative and quantitative data are collected/generated, processed and handled. Statistical data are needed from the involved end-users to get information on the current capacity (e.g. how many trains are on the track, how many passengers are transported, what is their average time on the track, at how many stations do the trains stop, etc). In addition, the project relies on meta-data about data sources to adapt interfaces, develop adapters to data sources and make the data accessible to the involved tools. In contrast to the meta-data the involved tools will have the actual data. Two kinds of durability of data are expected:

- Persistent data that will be used during the development and research work in work packages 3 to 7 to enhance algorithms and tools.
- Volatile data that will come into play especially in WP8 during simulation exercises in which the SAFETY4RAILS Information System (S4RIS) will be evaluated with real and/or simulated data.

This data will be used in SAFETY4RAILS to further develop the 18 involved tools as a core within the SAFETY4RAILS Information System (SRIS) to match to the requirements of the involved end users.

Overall, the type and formats of the data collected/generated in SAFETY4RAILS will be into three categories, identified to date, namely,

- Data collected/generated by direct input methods.
- Data from testing and training with SAFETY4RAILS developments.
- Data collected/generated from dissemination, communication and stakeholder engagement activities

By virtue of being a security project, the partners recognise that some of the data to be collected/generated may have a security connotation. It is not expected for the project to work with data with official "classified data" i.e. data requiring a Personal Security Clearance (PSC) certificate or Facility Security Clearance (FSC) certificate. The project's approach is also to avoid the use of such data. The project coordinator and Security Advisory Board will monitor and, if necessary, review this approach with recommendations to the Project General Assembly (PGA).

Ethically, any personal data collected/generated in the framework of SAFETY4RAILS is processed according to the principles laid out by the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data which entered into force in May 2018 aiming to protect individuals' rights and freedoms in relation to the processing of their personal data, while also facilitating the free flow of such data within the European Union.

Although SAFETY4RAILS is in its first six months, this report has endeavoured, as far as possible, to provide a framework which will ensure that both sensitive and non-sensitive data is properly managed, fully considering security and ethics.

8.2 Future work

This deliverable D9.5 represents information within the first six months of the SAFETY4RAILS project. It is therefore expected that further data will be generated for the purpose of the DMP. As such any updates will be incorporated in the subsequent deliverables D9.6 and D9.7.

Bibliography

Badii A, et al (2021): Deliverable D9.1 SAFETY4RAILS Ethical Compliance Framework (ECF). EU SAFETY4RAILS Project. Grant Agreement number 883532.

CASSDE (2020): Anonymisation. Online. https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/5.-Protect/Anonymisation. Last accessed 18 November 2020.

European Commission, TEMPLATE HORIZO 2020 DATA MANAGEMENT PLAN (DMP), available at: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm, last accessed 4 November 2020.

IPCC (2006a): 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Vol. 1 General Guidance and Reporting, CHAPTER 6 Quality Assurance / Quality Control and Verification.

IPCC (2006b): 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Vol. 1 General Guidance and Reporting, CHAPTER 6 Quality Assurance / Quality Control and Verification.

ITPro (2020): Hackers hold Newcastle Uni student data to ransom. Online. https://www.itpro.co.uk/security/ransomware/357022/hackers-hold-newcastle-uni-student-data-to-ransom. Last accessed 18 November 2020.

Steele L and Orrell T (2017). The frontiers of data interoperability for sustainable development. Publish What You Fund and Development Initiatives.

Sugimoto, S., Baker, T., & Weibel, S. L. (2002). Dublin Core: Process and Principles. Lecture Notes in Computer Science Digital Libraries: People, Knowledge, and Technology, 25-35.

UK Data Service (2020): Organising Data. Online. https://www.ukdataservice.ac.uk/manage-data/format/organising. Last accessed 18 November 2020.

Unige (2020): H2020 - Open Research Data Pilot. https://www.unige.ch/researchdata/en/make-plan/all/dmp-h2020/. Last accessed 27/11/2020.

ANNEXES

ANNEX I. GLOSSARY AND ACRONYMS

TABLE 10 GLOSSARY AND ACRONYMS

Term	Definition/description
AB	Advisory Board
AL	Activity leader
AP	Action point
CCTV	Closed-Circuit Television
D	Deliverable
DC	Data controller
DM	Dissemination manager
DMP	Data Management Plan
DMS	Document Management System
DoA	Description of the Action (Annex 1 to the Grant Agreement)
DOI	Digital Object Identifier
EB	Ethical Board
EC	European Commission
EM	Ethics manager
EUB	End-user Board
EUC	End-users coordinator
EXM	Exploitation manager
FAIR	Findable, Accessible, Interoperable and Re-usable
FP	Framework Programme
GDPR	General Data Protection Regulation
IM	Innovation manager
IPR	Intellectual Property Rights
MIN	Minutes
ORDP	Open Research Data Pilot
PC	Project coordinator
PFR	Periodic financial report

PGA	Project General Assembly
PID	Persistent Identifier
PMB	Project Management Board
PMT	Project Management Team
PR	Partner representatives
PRES	Presentation
PTR	Periodic technical report
PU	Public
QA	Quality Assurance
QC	Quality Control
QM	Quality manager
REA	Research Executive Agency
RPT	Report
S4RIS	SAFETY4RAILS Information System
SAB	Security Advisory Board
SM	Standardisation manager
SR	Semestral report
T	Task
TL	Task leader
TM	Technical manager
ToC	Table of Contents
TRL	Technology Readiness Level
WP	Work package
WPL	Work package leader
WTL	Work package Task Leader

SAFETY4RAILS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 883532.