



Funded by  
the European Union



## MYRTUS

Multi-layer 360° dYnamic orchestration and interopeRable design environmenT for compute-continUum Systems

Deliverable:

D12.1- Data Management Plan v1

Due date of deliverable: (30-06-2024)

Actual submission date: (01-07-2024)

Start date of Project: 01 January 2024

Duration: 36 months

Responsible: TU Dresden

Revision: draft

Dissemination level		
PU		
DEC		
SEN	Sensitive	x



## DOCUMENT INFO

### Author

Author	Company	E-mail
Jeronimo Castrillon	TUD	jeronimo.castrillon@tu-dresden.de
Francesca Palumbo	UNICA	francesca.palumbo@unica.it
Danilo Pani	UNICA	danilo.pani@unica.it
Tiziana Fanni	ABI	tiziana.fanni@abinsula.com
Katiuscia Zedda	ABI	katiuscia.zedda@abinsula.com

### Document history

Document version #	Date	Change
0.1	16.05.2024	Initial structure
0.2	07.06.2024	Tables describing data
0.3	07.06.2024	Modified table of contents (section reshuffling) and partners contribution
0.4	12.06.2024	Reorganized tables, created separate tables for results.
0.5	17.06.2024	Contribution integrated and document ready for internal review.
0.6	27.06.2024	Receive and integrate revisions.
0.7	28.06.2024	Document ready for final submission

### Document data

Keywords	Data management, findable, accessible, interoperable, reusable, repository, open source
Editor Address data	Name: Jeronimo Castrillon Partner: TU Dresden Address: Helmholtzstrasse 18, 01069 Dresden, Germany Phone: +49 351 -463 427 16

*Funded by the European Union, by grant No. 101135183. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.*



## Table of Contents

1	Executive Summary	4
1.1	Structure of the document	4
2	Data summary	5
3	Other research outputs	11
4	FAIR data	17
4.1	ABINSULA SRL	18
4.2	UNIVERSITA' DEGLI STUDI DI SASSARI	19
4.3	SOFTEAM	19
4.4	TECHNISCHE UNIVERSITAET DRESDEN	20
4.5	NEDERLANDSE ORGANISATIE VOOR TOEGEPAST NATUURWETENSCHAPPELIJK ONDERZOEK	20
4.6	UNIVERSITA' DEGLI STUDI DI CAGLIARI	21
4.7	LAKESIDE LABS GMBH	22
4.8	UNIVERSIDAD POLITÉCNICA DE MADRID	22
4.9	HIRO MICRODATACENTRES B.V	23
4.10	KING'S COLLEGE LONDON	24
4.11	Forge Reply S.r.l. con socio unico	24
4.12	ARUBAKUBE SRL	24
4.13	CANON RESEARCH CENTRE FRANCE	24
4.14	UNIVERSITA' DELLA SVIZZERA ITALIANA	25
5	Allocation of resources	25
6	Data security	25
7	Ethics	26
8	Other issues	27



## 1 Executive Summary

This document describes the current status of the Data management plan (DMP) of the MYRTUS project at M6. It is based on an early assessment of the characteristics of the data that will be used and generated in MYRTUS and how they adhere to the FAIR principles. We also provide an outlook on other project results, with a particular focus on software artefacts. Moreover, we detail the strategies of the institutions of the MYRTUS consortium with respect to FAIR practices, what resources are allocated to ensure the DMP, and describe aspects of data security and ethics. This is the first release of the plan, which will be updated at M22.

### 1.1 Structure of the document

The data management plan includes a summary of data in Section 2 followed by a summary of other research outputs in Section 3. Section 4 describes how the partners of MYRTUS adhere to the FAIR principles. Sections 5-7 are concerned with resource allocation, data security and ethical concerns, respectively. Finally, Section 8 includes miscellaneous aspects of data management not covered in the previous sections.



## 2 Data summary

This section describes an initial assessment of the data used and generated in the MYRTUS project. For every data item, Table 1 details the role of the data within the MYRTUS project and its relation to the project objectives, whether the data is re-used from existing data sources or is generated during the project, what formats will be used, how it will be stored, made available and what expected size the data will have. Every Myrtus Data (MD) item receives a unique identifier, like MD1, for further referencing in other documents of the project. Since MYRTUS has a strong emphasis on software components, code plays a key role within the project. How code is managed is discussed in Section 3.

*Table 1 MYRTUS data*

ID	Item	Description
MD1	Dataset name and reference	MYRTUS requirements set
	Dataset description	The requirements have been carefully defined to address all MYRTUS operational objectives and technical discussions. [Generic requirements] This set of requirements is intended to drive Pillar1, Pillar 2 and Pillar 3 development, respectively, in WP3-4, WP5-6, and WP7-8. [Specific requirements] Set of specific technical requirements necessary to implement the MYRTUS infrastructure and enable the MYRTUS computing continuum.
	Format	.csv file
	Expected size	KB
	Provenance	Generated within WP1-2 activities
	Is dataset confidential? Must be encrypted?	No
	Data sharing/access inside MYRTUS	[Generic requirements] Public [Specific requirements] Public
	Data sharing/access outside MYRTUS for research	[Generic requirements] Public [Specific requirements] Confidential
	Is dataset reusable?	Yes
	Envisioned Archive and Publishing Repository	[Generic requirements] Zenodo [Specific requirements] To be defined
	Responsible partner	UNICA-ABI
MD2	Dataset name and reference	Dataset name : FMDC and Edge Device Data Streams Reference : This dataset includes various streams of data and metrics collected from FMDC and edge devices, which are part of a distributed knowledge base extending to the cloud and edge environments.



	Dataset description	The dataset comprises data and metrics from FMDCs and edge devices that monitor and manage performance, operational status, and other relevant parameters. The data is processed at the fog layer, ensuring efficient data handling and reduced latency. The dataset supports real-time analytics and decision-making processes.
	Format	Dataset is available in common data formats such as CSV, JSON, and Parquet to facilitate ease of use and integration.
	Expected size	KB-MB
	Provenance	The dataset is collected from FMDCs and edge devices, processed to enrich the raw data into actionable insights, and stored in a distributed knowledge base. The data undergoes multiple stages of processing, including initial collection, fog layer processing, and final storage.
	Is dataset confidential? Must be encrypted?	No
	Data sharing/access inside MYRTUS	Public
	Data sharing/access outside MYRTUS for research	Public
	Is the dataset reusable?	Yes
	Envisioned Archive and Publishing Repository	Github
	Responsible partner	HIRO
<b>MD3</b>	Dataset name and reference	Emergent Scheduler (synthetically produced dataset)
	Dataset description	Dataset synthetically produced that presents a principle e-f-c infrastructure (no. of nodes on each layer, node connections inter and intra layer) and the incoming pods that need to be processed (incl. CPU/RAM demand, time sensitivity and privacy information). This data set will serve as input mainly to MO5 (Section 3 of this document).
	Format	CSV/JAML files
	Expected size	KB
	Provenance	Produced by LAKE during the project
	Is the dataset confidential? Must be encrypted?	No
	Data sharing/access inside MYRTUS	Public
	Data sharing/access	Public



	outside MYRTUS for research	
	Is the dataset reusable?	These data will be allowed for reuse following standard digital practices by naming the reference (preferable by a paper citation or cross-link to the git repo or arXiv).
	Envisioned Archive and Publishing Repository	GitHub/arXiv
	Responsible partner(s)	LAKE
<b>MD4</b>	Dataset name and reference	High-level application models
	Dataset description	Models or specifications by means of DSLs to describe the logic of applications components or kernels. Needed by the DPE to integrate in the code generation.
	Format	Source code of open benchmarks (e.g., from TensorFlow or other Python frameworks), or from DSLs (e.g., Halide, Cfdlang, LinguaFranca)
	Expected size	KB-MB
	Provenance	From open source benchmarks and open benchmarks produced by partners
	Is dataset confidential? Must be encrypted?	No
	Data sharing/access inside MYRTUS	Public Access
	Data sharing/access outside MYRTUS for research	Public Access
	Is dataset reusable?	Yes
	Envisioned Archive and Publishing Repository	GitHub/Gitlab and similar source code repositories
	Responsible partner	TUD, UNICA, UPM, UNISS
<b>MD5</b>	Dataset name and reference	Application evaluation results
	Dataset description	Results obtained by simulation (Mocasin) or executions in HW prototypes for the Node-level deployment. Needed to report results of the optimization framework of the DPE. This includes performance, area, and energy numbers as well as traces of execution in some cases.
	Format	csv files and documentation. In some cases, Artifacts are submitted along with scientific publication. For traces of execution, we will use Mocasin trace open format. Other formats may be considered in the project.
	Expected size	MB-GB



	Provenance	Generated by tools of the DPE and/or HW (e.g., Moccasin)
	Is dataset confidential? Must be encrypted?	No
	Data sharing/access inside MYRTUS	Public access
	Data sharing/access outside MYRTUS for research	Public access
	Is dataset reusable?	Yes (can be replotted)
	Envisioned Archive and Publishing Repository	GitHub/Gitlab and similar source code repositories
	Responsible partner	TUD
<b>MD6</b>	Dataset name and reference	Meta data of deployment variants
	Dataset description	Includes information of different implementation variants or operating points generated by the DPE to be leveraged at runtime. This information is key to aid the MIRTO cognitive engine for runtime decisions. Along with a given application implementation, the information will contain details about possible knobs (e.g., number of threads, mapping to dedicated HW if available or configuration of reconfigurable hardware) and the corresponding execution metrics that the engine may expect (e.g., latency, energy and area).
	Format	To be defined in the project
	Expected size	KB
	Provenance	Generated by the DPE
	Is dataset confidential? Must be encrypted?	No
	Data sharing/access inside MYRTUS	Public Access
	Data sharing/access outside MYRTUS for research	Public Access
	Is dataset reusable?	Yes (can be interpreted by other runtime engines)
	Envisioned Archive and Publishing Repository	GitHub/Gitlab and similar source code repositories
	Responsible partner	TUD
<b>MD7</b>	Dataset name and reference	Machine Learning evaluation data





	Dataset description	Defines the application scenarios for evaluation consisting of machine learning tasks and their requirements (quality, latency, and throughput). Standard benchmarks from industry and academia (i.e., MLPerf benchmark suite) will serve as early evaluation and will make it possible to compare Myrtus against the state-of-the-art. Applications may cover vision tasks like image classification and object detection, audio tasks such as keyword spotting and anomaly detection, and language processing tasks.
	Format	Specific to each machine learning task (JPG images, text, audio files).
	Expected size	GB
	Provenance	Open benchmark suite (MLPerf <a href="https://mlcommons.org/benchmarks/">https://mlcommons.org/benchmarks/</a> )
	Is dataset confidential? Must be encrypted?	No
	Data sharing/access inside MYRTUS	Public Access
	Data sharing/access outside MYRTUS for research	Public Access
	Is dataset reusable?	Yes
	Envisioned Archive and Publishing Repository	Original benchmark repositories
	Responsible partner	TUD
<b>MD8</b>	Dataset name and reference	Mobility UC images
	Dataset description	Blurred images captured by cameras, as input to road user detection process, for test/debug purpose.
	Format	Jpeg or h264 files
	Expected size	MB-GB
	Provenance	Captured by cameras
	Is dataset confidential? Must be encrypted?	Confidential, due to privacy management rules
	Data sharing/access inside MYRTUS	No sharing, access by CRF
	Data sharing/access outside MYRTUS for research	No sharing outside MYRTUS
	Is dataset reusable?	No



	Envisioned Archive and Publishing Repository	Only temporarily stored, for debug purpose, within CRF premise
	Responsible partner	CRF
<b>MD9</b>	Dataset name and reference	Mobility UC Edge and fog device related MIRTO metrics
	Dataset description	Metrics generated by the cameras and fog server, to serve as input to MIRTO cognitive engine.
	Format	Csv or json
	Expected size	MB
	Provenance	Generated by edge and fog devices
	Is dataset confidential? Must be encrypted?	No
	Data sharing/access inside MYRTUS	Public Access
	Data sharing/access outside MYRTUS for research	Public Access
	Is dataset reusable?	Yes (can be interpreted by other runtime engines)
	Envisioned Archive and Publishing Repository	MYRTUS Zenodo community
	Responsible partner	CRF, TNO ?
<b>MD10</b>	Dataset name and reference	Mobility UC Generated metadata (object lists) and ITS messages
	Dataset description	Metadata and ITS messages payloads generated from camera image processing and vehicle. Generated ITS messages might be CPM, CAM, SPAT, or MAP, depending on the scenario.
	Format	Json, or yaml
	Expected size	MB
	Provenance	Edge and fog Myrtus devices
	Is dataset confidential? Must be encrypted?	No
	Data sharing/access inside MYRTUS	Public Access
	Data sharing/access outside MYRTUS for research	To be defined
	Is dataset reusable?	yes



	Envisioned Archive and Publishing Repository	MYRTUS Zenodo community
	Responsible partner	CRF, TNO
<b>MD11</b>	Dataset name and reference	BSCR - Biomechanical signals from individual and collaborative rehabilitation tasks
	Dataset description	The dataset will include the information associated with the participants' movements during the exergame execution, including kinematic and electrophysiological data. The full data will be provided for a single participant exercising alone or with virtual agents only, or for multiple participants exercising together. The information will be correlated to the exercise experience.
	Format	Header file with metadata (.txt) Binary file for every participant in every single trial (.bin) Reader files for MATLAB will be provided.
	Expected size	MB
	Provenance	Edge and fog Myrtus devices
	Is dataset confidential? Must be encrypted?	No
	Data sharing/access inside MYRTUS	Public Access
	Data sharing/access outside MYRTUS for research	Public access
	Is dataset reusable?	yes
	Envisioned Archive and Publishing Repository	MYRTUS Zenodo community
	Responsible partner	UNICA

### 3 Other research outputs

In addition to the management of data, described in the previous section, MYRTUS will manage other research outputs, including software and demonstrators. These kinds of outputs are different research outputs with respect to data, and the FAIR principles are not directly applicable to their entirety. However, recently an effort has been made to extend the FAIR principles to them, for instance the FAIR4RS (FAIR for Research Software) has defined a list of recommendations<sup>1</sup> that apply to software that will be kept into account in MYRTUS. We thus include an initial assessment of code and demonstrators with a characterization aligned with the data summary in Section 2. Myrtus Outputs (MO) are described in Table 2.

<sup>1</sup> <https://doi.org/10.15497/RDA00068>



Table 2 MYRTUS research outputs

ID	Item	Description
MO1	Output name and reference	Communication Dissemination and Exploitation results
	Output description	Code and artifacts, including generated demonstrators for dissemination, communication and exploitation purposes. This includes running examples and tutorials on specific technologies.
	Format	We expect formats like .xdf, .c, .cpp, .pdf
	Expected size	The size of these outputs is not predictable at this moment.
	Provenance	Generated within WPs 11 and 12 implementation
	Output sharing/access inside MYRTUS	Public
	Output sharing/access outside MYRTUS for research	Public
	Is this output reusable?	Outputs will be made reusable following standard digital practices by naming the source of the information (e.g. link to repository with the artifact).
	Envisioned Archive and Publishing Repository	Zenodo or other open access repositories like GitHub
	Responsible partner	ALL
MO2	Output name and reference	Examples of CGR accelerators developed for R1
	Output description	Input specification, design steps, and output reference for reconfigurable accelerators included in the MYRTUS infrastructure
	Format	Source code (MDC-compliant xml or other dataflow spec. + Verilog or C++ for HLS), design steps (.md guide and .tcl or .sh scripts), output reference (verilog and block design).
	Expected size	KB - MB
	Provenance	Produced by partners during the project
	Output sharing/access inside MYRTUS	Public
	Output sharing/access outside MYRTUS for research	Public
	Is Output reusable?	Yes
	Envisioned Archive and Publishing Repository	Github
	Responsible partner	UNICA, UNISS



<b>MO3</b>	Output name and reference	Instances of the Multi-grain CGR Array developed for R2
	Output description	Specifications, list of prerequisites and tools, design methodology, interface definition, source code, output implementation to be included in the MYRTUS infrastructure, and user guide.
	Format	Source code (SystemVerilog + c for low-level drivers), specifications, design steps, and interface definition (.md guide and .tcl scripts), output reference (netlist exchangeable format).
	Expected size	KB-MB
	Provenance	Produced by UPM during the project
	Output sharing/access inside MYRTUS	Public
	Output sharing/access outside MYRTUS for research	Public
	Is the Output reusable?	Yes
	Envisioned Archive and Publishing Repository	GitHub/Gitlab and similar source code repositories
	Responsible partner	UPM
<b>MO4</b>	Output name and reference	Source code of the LLM workload optimization model
	Output description	Software developed and used for the training, testing and validation of the LLM model. The aim of the model is to optimize, according to specific criteria (energy, security, latency, ecc.), the code of the application (workload) that will run on the MYRTUS infrastructure.
	Format	Source code written mainly in python. Jupyter notebook files.
	Expected size	KB - MB
	Provenance	Produced by partners during the project
	Output sharing/access inside MYRTUS	Public
	Output sharing/access outside MYRTUS for research	Public
	Is the Output reusable?	This code will be allowed for reuse following standard digital practices by naming the source of the information (e.g. add the paper citation or cross-link to the website material).
	Envisioned Archive and Publishing Repository	GitHub



	Responsible partner(s)	UNISS
<b>MO5</b>	Output name and reference	Emergent Scheduler (Source code)
	Output description	Software framework developed and used for the emergent scheduling in a 360° e-f-c (related to MD5). The aim of the framework is to develop reactive swarm algorithms for the future dynamic MYRTUS infrastructure.
	Format	Source code written mainly in python, strongly utilizing the Python MESA library
	Expected size	KB - MB
	Provenance	Produced by LAKE during the project
	Output sharing/access inside MYRTUS	Public
	Output sharing/access outside MYRTUS for research	Public
	Is the Output reusable?	The code will be allowed for reuse following standard digital practices by naming the reference.
	Envisioned Archive and Publishing Repository	GitHub
	Responsible partner(s)	LAKE
<b>MO6</b>	Output name and reference	FREVO interface extension to DynAA simulator <a href="https://frevo.sourceforge.net/">https://frevo.sourceforge.net/</a> Source code
	Output description	FREVO is a software framework for producing the local rules for the 360° MYRTUS agents using evolutionary algorithms. Throughout the project we develop a interface to DynAA as the framework requires a runnable simulation environment.
	Format	Source code written mainly in JAVA
	Expected size	KB - MB
	Provenance	Produced by LAKE during the project; interface with the support of TNO
	Output sharing/access inside MYRTUS	Public
	Output sharing/access outside MYRTUS for research	Public
	Is the Output reusable?	These code will be allowed for reuse following standard digital practices by naming the reference (preferable by a paper citation or cross-link to the git repo).



	Envisioned Archive and Publishing Repository	GitHub or sourceforge
	Responsible partner(s)	LAKE, TNO
<b>MO7</b>	Output name and reference	Yocto-based OS Generation Files
	Output description	Generation, with the Yocto Project strategy, of custom distributions of Linux-based OS.
	Format	Source code in terms of configuration files (.conf) and BitBake recipes (.bb).
	Expected size	KB - MB
	Provenance	Produced by partners during the project
	Output sharing/access inside MYRTUS	Public
	Output sharing/access outside MYRTUS for research	Public
	Is the Output reusable?	Yes
	Envisioned Archive and Publishing Repository	GitHub, GitLab or sourceforge
	Responsible partner(s)	UNISS
<b>MO8</b>	Output name and reference	Intermediate code
	Output description	Intermediate code needed to explain and demonstrate transformations within the DPE, useful to the community to reproduce and gain insight from the technology developed within MYRTUS.
	Format	Intermediate code using MLIR dialects and/or LLVM
	Expected size	KB-MB
	Provenance	Generated by the DPE
	Output sharing/access inside MYRTUS	Public access
	Output sharing/access outside MYRTUS for research	Public access
	Is Output reusable?	Yes
	Envisioned Archive and Publishing Repository	GitHub/Gitlab and similar source code repositories
	Responsible partner	TUD, UPM, UNICA



<b>MO9</b>	Output name and reference	MLIR dialects
	Output description	New dialects (e.g., for improved dataflow support, for reactive programs, for CGRA modeling) in the MLIR ecosystem along with possible extensions to existing dialects.
	Format	C++ and Python code defining MLIR dialects
	Expected size	KB-MB
	Provenance	Developed in WP7-8
	Output sharing/access inside MYRTUS	Public access
	Output sharing/access outside MYRTUS for research	Public access
	Is Output reusable?	Yes
	Envisioned Archive and Publishing Repository	GitHub/Gitlab and similar source code repositories
	Responsible partner	TUD, UPM, UNICA
<b>MO9</b>	Output name and reference	MYRTUS Modelio Attack Trees Library [a link to the library on Zenodo <sup>2</sup> will be shared at M34]
	Output description	Collection of attack trees diagrams showing how MYRTUS Use case assets or a target might be attacked. These diagrams are intended for security specialists for modelling the attacks that occur on IT systems or cyber-physical systems and describe the events that lead to the attack in the form of a tree in which the attack is represented by the root element which is related with “OR” and “AND” conditions to the children that represent the sequence of events that lead to the root attack
	Format	pdf
	Expected size	MB
	Provenance	Use Case Analysis
	Output sharing/access inside MYRTUS	Public Access - The Myrtus Attack Trees Library will be public (only information that can be public will be added)
	Output sharing/access outside MYRTUS for research	Public Access - The Myrtus Attack Trees Library will be public (only information that can be public will be added)
	Is Output reusable?	Yes, with metadata including security attributes for the attacks such as the severity and the likelihood of the attack
	Envisioned Archive and Publishing Repository	MYRTUS Private Shared Google Drive & MYRTUS public web site & MYRTUS Zenodo website & MYRTUS social media

<sup>2</sup> <https://zenodo.org/communities/myrtus/records?q=&l=list&p=1&s=10&sort=newest>





	Responsible partner	SOFT, USI
<b>MO10</b>	Output name and reference	MYRTUS Model of the privacy preserving primitives [a link to the library on Zenodo <sup>3</sup> will be shared at M34]
	Output description	Collections of models of privacy preserving primitives that will be used by the security and privacy manager in MYRTUS.
	Format	pdf
	Expected size	MB
	Provenance	Use Case Analysis
	Output sharing/access inside MYRTUS	Public Access
	Output sharing/access outside MYRTUS for research	Public Access
	Is Output reusable?	Yes
	Envisioned Archive and Publishing Repository	MYRTUS Private Shared Google Drive & MYRTUS public web site & MYRTUS Zenodo website & MYRTUS social media
	Responsible partner	USI
<b>MO11</b>	Output name and reference	Simulation code of the proposed AFL scheme.
	Output description	Source code developed for the training, testing of proposed asynchronous federated learning scheme.
	Format	Source code written mainly in Python.
	Expected size	MB
	Provenance	Produced by KCL during the project.
	Output sharing/access inside MYRTUS	Public
	Output sharing/access outside MYRTUS for research	Public
	Is Output reusable?	This code will be allowed for reuse following standard digital practices by naming the source of the information (e.g. add the paper citation or cross-link to the website material);
	Envisioned Archive and Publishing Repository	GitHub
	Responsible partner	KCL

#### 4 FAIR data

MYRTUS fosters contributions to open-source SW and HW, exploiting both open results from previous/on-going partners' projects and collaborations, and from the open communities,

<sup>3</sup> <https://zenodo.org/communities/myrtus/records?q=&l=list&p=1&s=10&sort=newest>



aiming at their growth and integration with MYRTUS results and tools. MYRTUS approach is based on early release of open outputs, that are progressively updated during project lifetime.

This section describes provisions from each partner to ensure that data are accessible to research, education, businesses and governments across the EU in a way that meets European values and requirements. In this context, MYRTUS is aligned to the Open Science principles promoted by the European Commission, improving the diffusion of generated knowledge and data, and fostering open cooperative work, maintaining a solid adherence to the FAIR principles. In general MYRTUS partners will aligned themselves to the general FAIR strategy of MYRTUS:

- *Findable*: MYRTUS (meta)data and publications are shared through databases that provide persistent identifiers (e.g., DOI and URN).
- *Accessible*: MYRTUS adheres to the ‘as open as possible, as closed as necessary’ principle. MYRTUS (meta)data will be made open responsibly, without any sensitive information that could lead to the privacy of possibly involved people and according to the MYRTUS IPR strategy and Data Management Plan.
- *Interoperable*: interoperability is one of MYRTUS key selling strengths, to lower vendor/platform lock-in barriers. Similarly, (meta)data are going to be stored in open and usable file formats, using the representation most adopted in the related application fields.
- *Reusable*: MYRTUS (meta)data are released with clear and accessible documentation and data usage licence. Scientific publication will follow the Open Access guidelines for Horizon Europe, to ensure transparency of the processes and outcomes of research and disseminate new knowledge to foster open science.

Each partner is responsible for the curation and storage of their data in their servers, guaranteeing proper management and accessibility according to the applicable legislations and GDPR and in line with the MYRTUS FAIR strategy. Each partner will decline this strategy, according to the resources they have available, to manage research data. Please, notice that as mentioned in Section 3, there are other outputs for which the FAIR principles are nor directly applicable, but they will be made as FAIR as possible depending on the kind of output.

Please, note that the individual strategies have been communicated internally to the project and revised, to ensure they are aligned with the overall MYRTUS FAIR strategy, and approved. Activities related to data will be monitored and managed by the Project Coordinator with the support of the WP leaders and corrective actions will be put in place whenever is necessary. The following subsections detail the plans from each partner in MYRTUS.

#### 4.1 ABINSULA SRL

Abinsula research outputs such as publications, presentations and promotional materials will be made available in compliance with the general FAIR MYRTUS, in particular:

- *Findable*: these data will be uploaded to the MYRTUS Zenodo community, which provides a DOI for each contribution.
- *Accessible*: these data will be made available open source.
- *Interoperable*: for each author that has an ORCID, the related ID will be stated in the publication and in the publication metadata.
- *Reusable*: data uploaded on Zenodo are associated with specific licences that clarify how they can be (re-)used.



Abinsula technical activities are related to the development of the Multi-Sensor Gateway. Abinsula does not own the data that pass through the Gateway (or are processed in it) and is not expected to contribute to create/feed a dataset. The code related to the Gateway is archived in the Abinsula Git (<https://gitlab.abinsula.com/>). As mentioned in Section 3, code itself is not required to be FAIR. ABI will consider the FAIR4RS recommendations for research software in the developments of its Gateway.

#### 4.2 UNIVERSITA' DEGLI STUDI DI SASSARI

The Institutional Archive of scientific and technical production of the University of Sassari is established in order to accommodate the contributions of the scientific and technical studies and research, through the Institutional Research Information System (IRIS), a single institutional archive, from which the components of the organisation draw according to their respective needs, also to comply with regulatory and National Evaluation Agency of the University System and Research (ANVUR) provisions. University policy respects the F.A.I.R. principles as shortly described below.

- *Findable*: Publications will be findable on UNISS IRIS institutional archive and Zenodo. The UNISS IRIS institutional archive is i) equipped with advanced functions for the registration, certification, and dissemination of contributions; ii) indexed by the main generalist and specialist search engines which guarantee maximum dissemination and visibility of the deposited materials; and iii) in line with international technical standards for the preservation of contributions over time, providing persistent identifiers such as DOI.
- *Accessible*: Publications will be accessible on UNISS IRIS institutional archive and Zenodo. Code will also be accessible on platforms such as GitHub or GitLab.
- *Interoperable*: UNISS IRIS institutional archive is interoperable with other institutional archives: the Italian Ministry of University and Research (MUR) databases. This also applies to the data archived in Zenodo.
- *Reusable*: When inserting the Contribution into the UNISS IRIS institutional archive, the Author provides the essential information of the agreement with the publisher (if not registered in ad hoc databases) and, if requested by the UNISS working group on open access and effective management of the institutional archive, a copy of the part of the contract which states the terms of the transfer of rights and the contractual conditions practised by the publisher, if any. In case of free access rights, the author is required to grant the University a non-exclusive, free, irrevocable and universal licence to publish his contribution respectively with free or free access in the Institutional archive. Following the granting of the licence, the University immediately publishes the licensed contribution for free or open access. If the University is the owner of the copyright on the contribution, upon acceptance of the contribution, it will immediately proceed to publish it with free access in the archive, without prejudice to cases of incompatibility with other rights and procedures.

#### 4.3 SOFTEAM

SOFT will actively foster FAIR principles in MYRTUS research and will contribute an Attack Tree Library dataset containing all the sharable attack trees developed with the Modelio Modeling Tool Attack Trees Designer Module during the project lifetime, in particular the ones related to the two project use cases. The F, A, I, R principles will be followed:

- *Findable*: the dataset will be findable and shared on the Zenodo project web site with a DOI.



- *Accessible*: the dataset will be accessible and shared on the Zenodo project web site with a DOI
- *Interoperable*: the dataset will foster interoperability, it will include Attack Trees Diagrams developed in the Open Source Modelio Attack Trees Designer Module, this module will offer an ergonomic environment for designers and will contain features that allow users to configure security attributes for the attacks such as the severity and the likelihood of the attack. Moreover it will allow referencing other trees, importing and exporting attack trees.
- *Reusable*: the dataset will foster reusability, users will be able to reference, import and export attack trees using the Open Source Modelio Attack Trees Designer Module.

#### 4.4 TECHNISCHE UNIVERSITAET DRESDEN

MYRTUS will benefit from the extensive research infrastructure at TUD and the associated institutes via the network of Dresden concept. TUD actively fosters FAIR principles in research, contributes to multiple NFDIs and provides sustainable and solid support for data management and related aspects such as data literacy in data analysis, data integration and data publication. At the intersection of all disciplines, topics and research areas of TU Dresden, the Center for Interdisciplinary Digital Sciences (CIDS) offers expertise to process scientific issues related to high-performance computing, data management, analysis, modelling and simulation, artificial intelligence, digital knowledge exchange, extraction and transfer, as well as digital tools in teaching, social interactions, and other digitization topics. The management of large amounts of data is possible by providing large data storage facilities. To complete the data lifecycle services, the ZIH provides tools for metadata management, archiving and data publication (institutional repository OPARA), digital documentation with electronic lab notebooks (pilot with data competence center Come2Data), as well as a research information system and an intranet/collaboration platform.

Based on the IaaS that ZIH provides, along with the MYRTUS Zenodo Community and the biggest open-source Community GitHub, TUD plans to follow the FAIR principles as following:

- *Findable*: All publications will be findable on TUD research information system as well as Zenodo with a DOI
- *Accessible*: All datasets will be accessible via TUD platform and Zenodo, open-sourced on GitHub as well
- *Interoperable*: The interoperability is enabled by the fruitful MLIR framework, which is known as its highly-customizable feature to implement one's own DSL (Domain Specific Language). In addition, open-source frameworks such as PyTorch, ONNX and FINN will also be leveraged in this scope.
- *Reusable*: All datasets can be reusable by following the license requirements. Users can reference them with no drawbacks.

#### 4.5 NEDERLANDSE ORGANISATIE VOOR TOEGEPAST NATUURWETENSCHAPPELIJK ONDERZOEK

TNO data such as publications, presentations, developed software and data will be made available in compliance with the general FAIR MYRTUS principals, in particular:

- *Findable*: data will be uploaded to the MYRTUS Zenodo community, which provides a DOI for each contribution;
- *Accessible*: data will be made available open source;
- *Interoperable*: for the provided software and data open industry standards will be used as much as possible in order to interoperate with the data and software;



- *Reusable*: a selection will be made of the usability and value of the data. Data uploaded on Zenodo will be associated with specific licences that clarify how they can be (re-)used.

As some of the data from the Mobility Use Case contains privacy sensitive data, not all data can be provided externally as open data. To be able to use the data within the project, TNO, CRF and the municipality of Helmond are creating a specific data usage contract it is specified what data TNO and CRF are allowed to use and process from the crossings in Helmond and how privacy should be ensured (like blurring and covering specific elements of the images/video streams). Faces and license plates for example need to be blurred and windows of buildings at crossing need to be covered so no images/videos data is collected from the window areas (but only data of the pedestrians, bicycles and cars on the street). Having a specific contract also means that not all data will be available for all partners within the MYRTUS project. Specific processing (like blurring and covering) of the data is required for some of the data (like the camera feeds/images) before it can be stored in TNO and CRF premises and possibly shared with internal or external parties (depending on contract details).

During the project TNO will (together with CRF) determine what data (software and collected and generated data) can be shared (also according to the contract with Helmond) and what data would be valuable for sharing with internal and external parties.

#### 4.6 UNIVERSITA' DEGLI STUDI DI CAGLIARI

The Open Access Institutional Archive of the University of Cagliari (UNICA IRIS) offers free access to the intellectual and scientific production of the institution. It is fully compliant with international standards (OAI-PMH) and the archived information can be found through specialized or generalist search engines such as Google and Google Scholar. The service also ensures the preservation of deposited documents over time.

Along with UNICA IRIS, UNICA will exploit the functionalities offered by Zenodo. This will grant the possibility to archive source code and open data. These will be archived in the MYRTUS Zenodo Community.

For the data collected in the framework of the healthcare UC, special care will be posed in protecting all the sensible information. In this regard, anonymization will be guaranteed by design, as the participants' identities will never be registered on the system. Data collected during the trials by the game server, will be sent to a storage server hosting the live dataset, and will be downloaded from it at the end of the pilots to be curated for the production of an open dataset.

For source code, the mdc-suite GitHub page will be used for code development. This will ensure code availability along the project development and ease the access and contributions from MYRTUS partners and other interested developers, researchers, and companies. Clear license and standard data formats are adopted in line with the recommendations of the FAIR4RS Principles.

Based on the described tools and infrastructure, UNICA will adhere to the FAIR principles:

- *Findable*: Data produced by UNICA will be grouped in the MYRTUS Zenodo Community to be associated with a permanent DOI. For the healthcare UC data, the production of Data Descriptors (i.e., scientific papers presenting the methodology behind data creation of the dataset) will be pursued. Such open-access publications increase the visibility and then findability of the dataset. Additionally publications will be findable on the UNICA IRIS institutional archive, which is indexed by major search engines; source code will be



findable on GitHub during development, which is the most used framework for open-source code.

- *Accessible*: Publications will be permanently accessible on UNICA IRIS and Zenodo. Code will also be accessible during its development on GitHub. For the healthcare UC data, the MYRTUS Zenodo Community will be used as an open repository.
- *Interoperable*: For publications, UNICA IRIS ensures interoperability among other institutional archives. This holds also for Zenodo archived data. The source code, according to FAIR4RS, will use standard formats for AI models (ONNX) and applications and transformations (MLIR). For the healthcare UC data, a single standard format cannot be defined because of the heterogeneity of the recorded signals in the pilots. Accordingly, binary files with disclosed encoding and ASCII text files will be created. To pursue interoperability, metadata (data explaining setup and measurements) will be written in English with IS measurement units and SNOMED terms for clinical health information. Code snippets to access the data in MATLAB will be also provided, along with readme files and pseudocode.
- *Reusable*: All data are reusable following licence requirements. Standard open-source licences will be used. For the healthcare UC data, Open Data Commons Attribution Licence v1.0 or similar/equivalent will be used.

#### 4.7 LAKESIDE LABS GMBH

Lakeside Labs is committed to FAIR principles and will therefore support MYRTUS' strategy in terms of these principles. In particular, the FAIR principles will be applied by Lakeside Labs as follows:

- *Findable*: the data, resulting methodologies and code will be findable on GitHub and/or sourceforge.
- *Accessible*: the data, resulting methodologies and code will be accessible and shared on GitHub and/or sourceforge.
- *Interoperable*: the source code will foster interoperability (providing interfaces for the emergent scheduler and working on an interface between FREVO and DynAA). The emergent scheduler will be developed in Python, FREVO in JAVA - both very common programming languages
- *Reusable*: the dataset as well as the code will foster reusability, users will be able to reference.

#### 4.8 UNIVERSIDAD POLITÉCNICA DE MADRID

UPM is committed to the FAIR principles. The UPM Digital Archive (<https://oa.upm.es/>) hosts academic and scientific documentation (theses, final year projects, articles, etc.) generated at the institution in digital format and makes it accessible via the Internet, in line with the Budapest Open Access Initiative and the Berlin Declaration, which the Universidad Politécnica de Madrid is a signatory of, and documents are retrievable from the main search engines.

For MYRTUS, the UPM Digital Archive will be used in parallel to the Zenodo repository, specially for publishing project results, such as bachelor and master theses, which are written in spanish, and so they have a more local scope.

Besides, UPM, together with other Spanish universities and the CSIC (Spanish Research and Science Council) have negotiated some "transformative agreements for access to electronic





information" with publishers to cover Article Processing Charges (APCs).<sup>4</sup> These will be available to UPM researchers in MYRTUS to foster open access publications, as envisaged for the project.

Based on these tools, UPM will apply the FAIR principles as follows:

- *Findable*: All publications will be findable on the UPM Digital Archive institutional archive and the MYRTUS Zenodo with a DOI.
- *Accessible*: Publications will be permanently accessible on the UPM Digital Archive and Zenodo. Research data will be accessible in the MYRTUS Zenodo Community. Code will also be accessible during its development on GitHub and/or Gitlab.
- *Interoperable*: For publications, the UPM Digital Archive ensures interoperability with other institutional archives. This also applies to data archived on Zenodo. The source code will use standard formats (C/C++), and exchangeable libraries (MLIR or ONNX)
- *Reusable*: All data and source code will be reusable according to licence requirements. Standard open-source licences will be used, and all the source code contributions will be accompanied by user guides and tutorials to make the adoption of the technical developments easier by third parties.

#### 4.9 HIRO MICRODATACENTRES B.V

Hiro will apply FAIR principles on the available open dataset.

- *Findable*: All source code will be findable on GitHub.
- *Accessible*: The accessibility principle guarantees that data can be accessed by interested parties through well-defined protocols. As outlined previously, all public data from the MYRTUS project will be archived in reputable repositories, aligning with the guidelines of the Horizon Europe programme.
  - **Project Private Repository (PrivRep)** - This Microsoft Teams/SharePoint platform is designated for the exchange of general documents, templates for internal collaboration, databases in the processing phase, and data restricted from public access, along with their metadata.
  - **Project Open Data Repository (DataRep)** - This repository is advised for the management and sharing of public data such as datasets, videos, images, presentations, and for referencing open publications and software repositories.
  - **Project Code Repository (CodeRep)** - This GitHub organisation is the recommended venue for handling software developed within the project, facilitating sharing among project participants and with the public.
- *Interoperable*: HIRO is committed to utilising open file formats as much as possible, along with standardised metadata for data description. HIRO is committed to the maxima of *as open as possible, as closed as necessary* to facilitate third-party access for the extraction, utilisation, replication, and dissemination of all publicly available datasets under Creative Commons Licences. Open-source software will be accessible via free software licences.
- *Reusable*: Reusability pertains to the capability for data to be efficiently repurposed in compliance with transparent and readily available data usage policies.

---

4

<https://www.upm.es/UPM/Biblioteca/ServiciosUsuario/ProfesoresInvestigadores?fmt=detail&id=CON07680>



#### 4.10 KING'S COLLEGE LONDON

KCL will actively foster FAIR principles in MYRTUS research and will contribute an Asynchronous Federated Learning code containing the training and testing of the AFL, evaluating the performance of the proposed schemes for AFL. The F, A, I, R principles will be followed:

- *Findable*: the data will be findable on GitHub.
- *Accessible*: the data will be findable and shared on GitHub.
- *Interoperable*: the dataset will foster interoperability, it will use the widely used datasets, i.e., MNIST and CIFAR10 for training. The code will be developed using the Pytorch framework, which is compatible for multiple platforms.
- *Reusable*: the dataset will foster reusability, users will be able to reference.

#### 4.11 Forge Reply S.r.l. con socio unico

Forge Reply will promote the FAIR principles in MYRTUS research and will contribute specifically in the “Healthcare Use Case”, developing the VR demonstrator. The FAIR principles to be followed are:

- *Findable*: All source code will be findable on GitHub.
- *Accessible*: All source code will be findable and shared on GitHub
- *Interoperable*: Datasets will be stored in open and accessible file formats (see UNICA section 4.6).
- *Reusable*: All data are reusable following licence requirements. Standard open-source licences will be used.

#### 4.12 ARUBAKUBE SRL

Arubakube data such as publications, presentations and promotional materials will be made available in compliance with the general FAIR MYRTUS, in particular:

- *Findable*: these data will be uploaded to the MYRTUS Zenodo community, which provides a DOI for each contribution.
- *Accessible*: these data will be made available open source.
- *Interoperable*: All data collected and generated will comply to cloud-native industry standards.
- *Reusable*: data uploaded on Zenodo are associated with specific licences that clarify how they can be (re-)used.

ArubaKube technical activities are related to the development of the Cloud-to-Edge Continuum. ArubaKube does not own the data that pass through the developed components (or are processed in it) and is not expected to contribute to create/feed a dataset. The code related to the continuum is archived in the ArubaKube's Git ( <https://github.com/ArubaKube>). As mentioned in Section 3, code itself is not required to be FAIR. ARK will consider the FAIR4RS recommendations for research software in the developments of its components.

#### 4.13 CANON RESEARCH CENTRE FRANCE

CRF will apply FAIR principles in MYRTUS research and will contribute to datasets generated during the project lifetime, in particular the ones related to the mobility use case. The F, A, I, R principles will be followed:





- *Findable*: the datasets will be findable and shared on the Zenodo project web site with a DOI.
- *Accessible*: the datasets will be findable and shared on the Zenodo project web site with a DOI
- *Interoperable*: Datasets will be stored in open and usable file formats, using the representation most adopted in the related application fields.
- *Reusable*: Data uploaded on Zenodo will be associated with specific licences and documentation, that clarify how they can be (re-)used.

#### 4.14 UNIVERSITA' DELLA SVIZZERA ITALIANA

USI is strongly committed to FAIR principles. This will also apply to the activities that will be carried out in the MYRTUS project. USI will contribute with SOFT to the attack tree development using Modelio and will work on the MYRTUS Privacy and Security Manager that will use the models of the privacy preserving libraries, also developed during MYRTUS. In details, the F, A, I, R principles will be achieved in this way:

- *Findable*: the dataset will be findable and shared on the Zenodo project repository with a DOI
- *Accessible*: the dataset will be accessible from the Zenodo project repository, format used to store the data will follow the practice of the target community and the software needed to access the file will be clearly listed and specified and will include an open source variant to ensure wider accessibility.
- *Interoperable*: the dataset and the accompanied metadata will follow best practice for interoperability. Where existing, standard format will be used for the data and metadata. All data and metadata will include the appropriate documentation to access and use the data and the metadata and to extend or import additional components (for instance, to import additional models of privacy preserving primitives).
- *Reusable*: the models will be reusable and will be possible to export existing models and to import new models to be used in the MYRTUS privacy and security manager.

#### 5 Allocation of resources

As mentioned above, each partner is responsible for the data management. The costs for making data and results FAIR are included in the budget of each institution. Apart from the detailed information provided in the previous sections, MYRTUS established a Zenodo community (referenced multiple times in this document). Open publications, presentations and so on will be uploaded to that repository. Partners missing an infrastructure to ensure data security can use the MYRTUS Zenodo community.

Apart from the above, partner **TUD** offers the OPARA service for archiving research data for at least ten years (in accordance with good scientific practice): <https://opara.zih.tu-dresden.de>. Data items are given a DOI for permanent referencing and stored indefinitely. OPARA allows storing data generated with the participation of the TUD, but does not have to be generated exclusively at the TUD. The upload is enabled for any TUD scientists or persons with a guest login. Prior to uploading an expert has to assess the quality of the data. OPARA has no restrictions regarding the maximum amount of data. Larger datasets that cannot be uploaded via the browser (e.g. > 5 GB) can be imported on the server by individual arrangement via the service desk without extra costs.

#### 6 Data security



Data security is of utmost importance in MYRTUS. We comply with local, EU, and international applicable legislations (such as the GDPR) and adhere to best practices with respect to security of data. Data platforms described throughout this document can be accessed only by authenticated users with proper access control. Access is granted via request to the coordinating team. Data maintained by partners themselves are securely stored, with remote access enabled via VPNs or encrypted communication. General security guidelines in MYRTUS include:

- Redundancy, e.g., geographical duplication, NAS and RAID drives.
- Device-level encryption should be the preferred choice.
- Limit the use of portable drives and USB sticks, while enforcing encryption of sensible data.
- Use state of the art access control policies and 2-factor authentication.

## 7 Ethics

In the context of Myrtus mobility Use Case, a set of cameras will be installed in two road intersections in the city of Helmond. In order to preserve people's privacy, cameras will be configured to output only metadata or blurred videos or images (face and licence plate blurring), and mask buildings as much as possible. Moreover, camera images will be processed on the fly and not stored on device by default. In case a few extracts need to be temporarily stored, for debug and/or test purpose for instance, they would be stored on a secured private storage, within CRF premise in France, with limited access, for a maximal duration of 6 months.

In the context of the healthcare UC, signals acquired from human subjects will be collected. These signals include surface electromyography (electrical activation of skeletal muscles, by wireless acquisition modules), heart rate, body segments orientation and kinematics (by inertial measurement units and/or controllers and trackers of a virtual reality headset), forces applied to a haptic joystick, etc. The data will be anonymously recorded on volunteers participating in the pilot tests. Pilots concurring to the dataset creation will be performed following the principles outlined in the Helsinki Declaration of 1975, as revised in 2000. All participants will sign an informed consent regarding data acquisition and use prior to their participation in the pilots. Data usage expressly includes signal analysis and processing for the extraction of the relevant information, machine learning to model participants' behaviour, anonymized open data sharing with the scientific community to foster the development of novel rehabilitation paradigms and technologies in the field. Data management will be performed according to the current 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, and repealing Directive 95/46/EC (General Data Protection Regulation). Data collected during the trials by the game server, will be sent to a storage server hosting the live dataset, and will be downloaded from it at the end of the pilots to be curated for the production of an open dataset. Servers will be hosted in the EU. The ethical approval by the unique Sardinian Ethics Committee, which is competent for this study, is currently being requested.

According to the GA, an independent Ethics Advisor was identified (see Deliverable 14.1) to verify the ethics of research and the adopted methodology, in particular with regards to the interaction of human subjects with AI tools: John H. Holmes, PhD, is Professor of Medical Informatics in Epidemiology at the University of Pennsylvania Perelman School of Medicine.



The EPQ - H - NEC - AI - POPD - Requirement No. 2 (deliverable D14.2) will be released by the end of July 2024, according to the schedule.

## 8 Other issues

Apart from all aspects, partners have several strategies to further support data management, as described in the following.

- TUD offers half-day and one-day workshops on RDM to researchers of the TUD. The courses give an introductory overview on the different aspects of RDM throughout the data life cycle and how it can be supported by tools and services. Topics include data storage, security, and flow, the description of data with metadata, personal and collaborative data management, archiving and the publication of research data to provide the participants with the knowledge to manage their data according to the FAIR principles. Extended workshops include hands-on experience.
- Through its International Doctoral School, UPM offers training courses related to data management for doctoral students. These courses include "Scientific Information: Research Support Resources" and "Data Management in Research Projects." Predoctoral researchers involved in MYRTUS will be invited to participate in these courses.