D1.3: Data Management Plan

Revision: v.3.0

Work package	WP 2
Task	Task 1.4
Due date	31/03/2026
Submission date	01/04/2026
Deliverable lead	NOVA
Version	3.0
Authors	Carla Ferreira (NOVA), João Leitão (NOVA),
Abstract	This document presents the final version of the data management plan, reflecting the complete data management practices adopted throughout the TaRDIS project.
Keywords	Data management, FAIR principles

Document Revision History

Version	Date	Description of change	List of contributor(s)
V1.0	20/03/2023	1st version of the data management plan	Carla Ferreira (NOVA)
V2.0	25/08/2024	2nd version of the data management plan	Carla Ferreira (NOVA)
V3.0	25/03/2026	Final version of the data management plan (project end)	Carla Ferreira (NOVA)

Grant Agreement No.: 101093006
Call: HORIZON-CL4-2022-DATA-01

Topic: HORIZON-CL4-2022-DATA-01-03
Type of action: HORIZON- RIA

DISCLAIMER

Funded by
the European Union

Funded by the European Union (TARDIS, 101093006). 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.

COPYRIGHT NOTICE

Project funded by the European Commission in the Horizon Europe Programme		
Nature of the deliverable:	**DMP**	
Dissemination Level		
PU	*Public, fully open, e.g. web (Deliverables flagged as public will be automatically published in CORDIS project's page)*	✓
SEN	*Sensitive, limited under the conditions of the Grant Agreement*	
Classified R-UE/ EU-R	*EU RESTRICTED under the Commission Decision No2015/ 444*	
Classified C-UE/ EU-C	*EU CONFIDENTIAL under the Commission Decision No2015/ 444*	
Classified S-UE/ EU-S	*EU SECRET under the Commission Decision No2015/ 444*	

* R: Document, report (excluding the periodic and final reports)

DEM: Demonstrator, pilot, prototype, plan designs

DEC: Websites, patents filing, press & media actions, videos, etc.

DATA: Data sets, microdata, etc.

DMP: Data management plan

ETHICS: Deliverables related to ethics issues.

SECURITY: Deliverables related to security issues

OTHER: Software, technical diagram, algorithms, models, etc.

EXECUTIVE SUMMARY

The following document is Deliverable 1.2 Data Management Plan (DMP) of the TaRDIS Project, funded by the European Union's Horizon Europe research and innovation programme under grant agreement Number 101093006.

This document is the final version of the DMP, providing a comprehensive account of the types and formats of data collected and generated throughout the project, their origins, data utility, and how TaRDIS's research data has been made findable, accessible, interoperable, and reusable (i.e., FAIR).

The purpose of the DMP is to provide the consortium members with an analysis of the main elements of the data management policy for all data generated by the project.

This deliverable has been maintained as a living document throughout the project, updated at regular intervals. This final version (V3.0) is submitted at month 39 and reflects the complete data management practices adopted during the TaRDIS project.

TABLE OF CONTENTS

Funded by the European Union

DEFINITIONS

Data: Observations or measurements (unprocessed or processed) represented as text, numbers, or multimedia.

Dataset: A structured collection of data.

Database: Datasets and other items stored together to serve one or more purposes or applications, often including data query or search and retrieval capabilities.

Metadata: A structured, machine-readable file that provides basic information about data (who, what, when, where, why, and how) that is essential to promote scientific collaboration; enable discovery, interpretation, and effective use of the data; and document its nature and quality.

Source Data: Primary or Secondary data used as input to produce products. Primary data is data measured or observed by the researcher, and is in a basic form that has been calibrated, converted to standard units, and has passed quality control procedures that remove or flag incorrect data. Secondary data is defined as data collected by someone other than the researcher.

Funded by the European Union

ABBREVIATIONS

API	Application Programming Interface
DMP	Data Management Plan
DPO	Data Protection Officer
EU	European Union
FAIR	Findability, Accessibility, Interoperability, and Reusability

Funded by the European Union

1 DATA SUMMARY

What is the purpose of the data collection/generation and its relation to the objectives of the project?

In TaRDIS there are different kinds of generated and collected data that directly contributed to the project objectives. First, data was collected from developers, software and network architects, and data scientists to identify their needs and determine the functional requirements of the TaRDIS development environment. This step was crucial in identifying and addressing the expectations of potential beneficiaries from various domains. Second, the main project results are a set of artefacts that comprise the TaRDIS toolbox. These artefacts include source code, algorithms, documentation, etc. Third, for the evaluation of the artefacts developed in the project, the consortium members employed synthesised data, either generated within TaRDIS or sourced from existing open source and (partner) proprietary datasets.
As stated in Section **5**, the data generation and collection complied with the EU ethics and legal requirements as well as national ethics and legal requirements.

What types and formats of data will the project generate/collect?

- Survey responses: text data in formats like CSV or JSON.
- Algorithm: stored as computer code or as documentation files.
- Computer code: stored as text files in code formats as Java, Python, among others.
- ML models: Saved in formats such as TensorFlow SavedModel or PyTorch .pth files.
- Equipment outputs: structured data files in formats like CSV or HDF5.
- Synthetic workloads: structured data files in formats like CSV or HDF5.
- Performance measurement logs: text data in formats like CSV or JSON.

Will the project re-use any existing data and how?

The project utilises existing code artefacts and algorithms, including programming models, APIs, the development environment, verification tools, partial replication datastore, swarm ML tools, and distributed runtime systems. While real data from the TaRDIS use cases is not currently available, it is expected to be incorporated as the project progresses. In the meantime, publicly available datasets have been used for ML modelling:

- Fashion-MNIST Dataset: Licensed under the MIT licence, this dataset evaluates personalised Federated Learning (FL) and FedAvg algorithm implementations within the Flower framework.
- MetroPT-3 Dataset: Utilised to assess the anomaly detection solution developed for the ACT use case.
- Solar Power Dataset: To support energy-related use cases, an LSTM model is trained using power values from solar panels in a smart home in Copenhagen, Denmark.
- CIFAR-10 Dataset via Microsoft's NNI Library: The Microsoft NNI library is used to test pruning techniques and early exit inference methods.
- Energy Data from Southern Germany: Data from small businesses and residential households in southern Germany is used to train a Deep Reinforcement Learning (DRL) model for the smart home energy management system.

These existing datasets provided a foundation for developing, testing, and validating algorithms and models within the TaRDIS project. They enabled accurate evaluation against real-world scenarios. As the project progressed, real data from the use case providers was also incorporated in the evaluation activities.

Additionally, the MNIST dataset was used for federated learning classification tasks (CNN model with FedAvg), and simulated ACT factory data (JSON logs of industrial process events from 17 emitters, totalling approximately 682 events) was used for anomaly detection

Funded by the European Union

evaluation. The project also generated use-case-specific ML models, including Physics-Informed Neural Networks (PINNs) for orbit determination in the GMV use case, and deep reinforcement learning models (DDPG and DQN) for smart home energy management in the EDP use case.

A significant output of the project is the set of open-source software tools developed under WP5 (distributed AI and AI-based orchestration). These tools, which constitute core components of the TaRDIS toolbox, include: the Flower-based FL tool (T-WP5-01/02/03) for federated learning model training, preprocessing, and inference; PTB-FLA and MPT-FLA (T-WP5-04) providing a Python testbed and MicroPython implementation for federated learning algorithms; the Fedra framework (T-WP5-09) for decentralised P2P federated learning; the FAuNO orchestrator (T-WP5-05) and PeersimGym environment for AI-driven task offloading; pruning (T-WP5-08), early exit (T-WP5-06), and knowledge distillation (T-WP5-07) tools for lightweight ML inference; and the DEXIT framework for decentralised early-exit inference deployment. All of these tools have been released under open-source licences via GitHub repositories (see Section 2.2 for details).

Work Package 6 (WP6) developed the foundational runtime, protocol, and data management substrate of TaRDIS. The core contribution is the Babel ecosystem, a protocol-centric runtime and programming model comprising Babel-Swarm for general-purpose swarm-scale deployments on the JVM, Babel-Android for mobile devices, and Micro-Babel (a C-based implementation on FreeRTOS) for resource-constrained embedded hardware such as ESP32 and RP2040 devices. On top of this runtime, WP6 delivered interoperable building blocks for decentralised membership management (HyParView, Cyclon, X-BOT, Random Tour), gossip-based message dissemination (eager gossip broadcast, flood broadcast, one-hop broadcast, and anti-entropy recovery), replicated data storage (Nimbus, a CRDT-based key-value store; PotionDB, a transactional store with materialised views; Arboreal, a hierarchical cloud-edge replication system with causal+ consistency; and a shared CRDT library), application-level streaming (decentralised pub-sub), telemetry and monitoring (Docker monitorisation framework with Prometheus and Grafana, Babel-Core metrics extension, decentralised telemetry aggregation), and secure coordination (decentralised energy market protocols). WP6 also developed adapters for external tools and libraries, including Cassandra, IBM Hyperledger Fabric, C3, Engage, and the Actyx middleware, as well as a REST/WebSocket API for language-agnostic access to the Babel runtime. The TaRDIS Messaging App demonstrator exercised the full WP6 protocol stack across servers, Raspberry Pi devices, Android phones, and ESP32-based IoT hardware. A large-scale experiment validated correct operation with 5,000 concurrent Babel-Swarm processes under emulated Internet-scale conditions, achieving 99.49% average reliability and approximately 1.2 seconds average delivery latency. The Babel Ecosystem has been identified as a Key Exploitable Result and is being pursued through the Horizon Results Booster programme for commercial exploitation. All WP6 software is hosted on the TaRDIS GitLab at NOVA (see Section 2.2 for repository details).

Work Package 4 (WP4) developed a comprehensive suite of formal analysis tools to ensure the soundness, security, and reliability of heterogeneous swarms. The final toolset covers communication behaviour analysis (WorkflowEditor and Actyx middleware for swarm protocol design, compositional verification of swarm protocols via Machine-runner and Machine-check, JoinActors for fair join pattern matching in Scala, Scribble for multiparty session type-based protocol verification, and JaTyC for Java typestate checking), data convergence and integrity (VeriFx for the design and verification of replicated data types with automated proofs, Ant for compile-time commutativity analysis to ensure data consistency, and AtomiS for data-centric synchronisation), security verification (IFChannel for information flow analysis over untrusted networks, PSPSP for protocol security verification in Isabelle/HOL, CryptoChoreo for choreographic specification of cryptographic protocols, and (Sec)ReGraDa-IFC for DCR choreography-based information flow control), and federated learning orchestration (correct

orchestration of PTB-FLA via CSP modelling and PAT verification, and correct hierarchical namespaces and dataspaces via session types and typed graph transformations). All WP4 tools have been integrated into the TaRDIS APIs, IDE, and AI optimisation framework developed under WP3. Key open-source repositories include Scribble (https://github.com/nuscr/nuscr), JoinActors (https://github.com/a-y-man/join-actors), and VeriFx (https://zenodo.org/records/7982416, with documentation on the TaRDIS wiki at https://codelab.fct.unl.pt/di/research/tardis/toolkit/Documentation/-/wikis/VeriFx).

What is the origin of the data?

Most of the code artefacts and algorithms that the project builds on top of were developed by the consortium partners, while a minority are (well-known) open-source tools. All credits regarding open-source tools (generated outside the consortium) have been clearly identified.

What is the expected size of the data?

Data originates primarily from consortium partners, supplemented by well-known open-source datasets. The estimated data sizes are:

- Survey responses: less than 10 MB
- Algorithms: less than 10 megabytes
- Computer code: tens to a few hundreds of MB (considering source code and compiled versions of computer code ready for execution across multiple platforms).
- ML models: less than 10 GB.
- Equipment outputs: less than 10 GB.
- Synthetic workloads: less than 10 GB.
- Performance measurement logs: Tens of GB.

To whom might it be useful ('data utility')?

The produced data has been useful to the TaRDIS consortium partners, and for dissemination, communication, and exploitation activities. Synthetic workloads can be of benefit for the scientific community at large, as it provides relevant benchmarks for some types of solutions and systems. Performance measurement logs can be useful also for the scientific community as to provide a basis for checking the reproducibility of results produced by the consortium.

Funded by the European Union

2 FAIR PRINCIPLES

2.1 FINDABLE PRINCIPLE, INCLUDING PROVISIONS FOR METADATA

Are the data produced and/or used in the project discoverable with metadata, identifiable, and locatable by means of a standard identification mechanism (e.g., persistent and unique identifiers such as Digital Object Identifiers)?

TaRDIS data has been shared between partners. Data was stored in files that followed a common template for all partners in the project: Reports, Presentations, Deliverables, and Milestones. To facilitate search, documents follow a specific naming and contain:

- a list of abbreviations used.
- a list of keywords.
- a list of the references used.

Previous files were not deleted; they were just saved in a folder labelled “Older Versions”.

Research data and outputs have been deposited and described in public data repositories that guarantee long-term preservation and assign persistent, unique identifiers to the deposited items. In particular, the TaRDIS Zenodo community (https://zenodo.org/communities/tardis-project/) assigns Digital Object Identifiers (DOIs) to all deposited items, ensuring that each dataset and output is uniquely and persistently identifiable. Zenodo also provides detailed metrics on downloads and views, enabling the consortium to assess the reach and impact of its outputs. All datasets have been made available with relevant metadata describing the data (data structure, units, conditions under which the data were generated or collected) to ensure their reusability.

The main source code repository is hosted on NOVA (GitLab), the partner responsible for ensuring compliance with relevant EU requirements and FAIR practices. Repositories in GitLab can either be private or public. Private repositories were used to share and support the development of solutions among consortium members. When an output was considered mature and had been evaluated, the repository was made available to the general community, with links to datasets related to the artefact in that repository (either input data used in testing or logs collected during the artefact's evaluation and validation).

What naming conventions do you follow?

The naming conventions followed are defined in deliverable D1.1 Project Plan.

Will search keywords be provided that optimise possibilities for re-use?

Yes, the TaRDIS project has provided a list of keywords for artefacts, data, publications, and other results. These have been defined and unified among the consortium partners. NOVA is responsible for verifying the proper use of those keywords in public repositories.

Do you provide clear version numbers?

Version numbers are provided in the file name and in the repository names.

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Metadata includes a collection of information, namely the title, author, description, file size and format (including units of any numerical data), type, publication date, keywords, access rights, licence, digital object identifier, related identifiers, and grant reference.

2.2 MAKING DATA OPENLY ACCESSIBLE

Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why clearly, separating legal and contractual reasons from voluntary restrictions.

Data produced and/or used in the TaRDIS project have been revised and approved by all consortium partners before being made openly available. This communication was sent to the Consortium and the Steering Committee with 30 days' notice to evaluate whether any disclosure of information or results could jeopardize patent prosecution under confidential information agreements.

All data published in scientific conferences or journals has been in line with the Horizon Europe guidelines and made available in open access. For this type of data, 15 days' notice was given by any partner wishing to use project-related data to that effect.

How will the data be made accessible (e.g., by deposition in a repository)? What methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g., in open-source code)?

Publications on project results and processes have been made available in open access, in accordance with a project-wise open-access publication strategy defined in the Consortium Agreement (CA). In particular, the consortium adopted the green open access model to make journal and conference publications publicly available. This implies that preprints or postprints of papers are available through well-known, highly reliable, perpetual archival sites, such as arXiv, under public copyright licences (e.g., "CC BY"). The original paper versions are freely available on the publisher's site for conferences and journals that offer the gold open-access model. In all cases, the project website provides details on how to retrieve official or pre- or post-print versions of scientific publications.

Regarding the algorithms, programming tools, ML models, and, more generally, the software developed during the project, the consortium used a repository (GitLab) to release all mature artefacts associated with the scientific publications under appropriate open-source licences. This strategy allows scientific communities to freely access the software used to derive the scientific results reported in publications, with acknowledgement of the use of TaRDIS resources. Access to this data is available via the git protocol (publicly available) or https.

In particular, the following WP5 software tools have been released as open-source repositories on GitHub: the Flower-based FL tool (https://github.com/lidijaf/Flower-based_FL_tool), PTB-FLA (https://github.com/miroslav-popovic/ptbfla), Fedra (https://github.com/anaskalt/fedra), FAuNO (https://anonymous.4open.science/r/FAuNO-C976/README.md), Pruning (https://github.com/Ilias-Paralikas/Pruning), Early Exit (https://github.com/Ilias-Paralikas/early_exit), Knowledge Distillation (https://github.com/levgiorg/KnowledgeDistillation), and DEXIT (https://github.com/anaskalt/dexit). Additionally, the Smart-Home-P2P-Energy-Trading-RL tool for the EDP use case was published at https://github.com/levgiorg/Smart-Home-P2P-Energy-Trading-RL. The WP5 tools are also documented on the TaRDIS wiki, under the Artificial Intelligence and Machine Learning APIs section (https://codelab.fct.unl.pt/di/research/tardis/toolkit/Documentation/-/wikis/TaRDIS-APIs/Artificial-Intelligence-and-Machine-Learning-APIs).

Over the course of the project, the partners produced a substantial body of scientific publications. WP5 alone contributed 15 conference papers and 9 journal articles spanning federated learning, reinforcement learning, lightweight ML techniques, and AI-driven orchestration. All publications follow the green or gold open-access model, with preprints or postprints deposited in arXiv and final versions available from conference or journal publishers.

The WP6 toolbox components, including the Babel ecosystem (Babel-Swarm, Babel-Android, and Micro-Babel), the communication and membership protocols (HyParView, Cyclon, gossip broadcast variants, anti-entropy), the replicated data stores (Nimbus, PotionDB, Arboreal), the external adapters (Cassandra, Hyperledger Fabric, C3, Engage, Actyx), and the telemetry and monitoring tools, are hosted on the TaRDIS GitLab at NOVA (https://codelab.fct.unl.pt/di/research/tardis). WP6 documentation, including API references and usage guides, is available on the TaRDIS wiki (https://codelab.fct.unl.pt/di/research/tardis/toolkit/Documentation/-/wikis/home). The oar-p2p experimental infrastructure used for the 5,000-node validation experiment is also available through NOVA's GitLab. WP6 contributed extensively to the scientific literature, with publications at venues including SRDS, ECOOP, PaPoC@EuroSys, and IEEE Access, all following the green or gold open-access model.

The WP4 formal analysis tools are distributed as open-source software through multiple platforms. Scribble, the multiparty session types toolchain for communication protocol verification, is available on GitHub (https://github.com/nuscr/nuscr) with a web-based interface at https://nuscr.dev/nuscr, and documentation on the TaRDIS wiki (https://codelab.fct.unl.pt/di/research/tardis/toolkit/Documentation/-/wikis/Scribble).
JoinActors, the Scala 3 library for fair join pattern matching, is available at https://github.com/a-y-man/join-actors. VeriFx, the framework for verified replicated data types, is deposited on Zenodo (https://zenodo.org/records/7982416) with documentation at https://codelab.fct.unl.pt/di/research/tardis/toolkit/Documentation/-/wikis/VeriFx. The WorkflowEditor and Actyx middleware tools (Machine-runner, Machine-check) are distributed as TypeScript libraries. The PSPSP security protocol verification tool is built on the Isabelle/HOL proof assistant. All WP4 tools are integrated into the TaRDIS APIs and IDE (WP3) via CI/CD pipelines. WP4 contributed publications at top venues including ECOOP, ESOP, POPL, ITP, and ICTAC, with preprints available on arXiv where applicable.

Synthetic workloads and performance logs used and collected during the evaluation of project results have been made available through a long-lived data repository hosted at NOVA, which provides an HTTPS interface for retrieving the data and metadata for those datasets.

Restrictions were applied only on account of privacy, ethical issues, copyright, confidentiality, and exploitation issues. Data that cannot be openly shared has been specified in the DMP. The DPO at NOVA (Dr João Marujo) addressed any specific questions or issues regarding compliance with GDPR and related legal implications.
All queries regarding scientific data storage and curation were the responsibility of the General Assembly.

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories that support open access where possible.

As mentioned in the previous item, research publications for conferences and journals that offer the gold open access model are available on the publisher's site. When the gold open-access model is not offered, preprint versions are available through perpetual archival sites such as arXiv. Source code and artefact distribution are available through the GitLab hosted

at NOVA (https://codelab.fct.unl.pt/di/research/tardis). Datasets are hosted on the TaRDIS GitLab platform to ensure long-term archival access to both datasets and metadata via an HTTPS interface. A Zenodo community specifically for the TaRDIS project has been established (https://zenodo.org/communities/tardis-project/). By creating a Zenodo community, TaRDIS can track the usage and impact of its datasets, as Zenodo provides detailed metrics on downloads and views. This data is invaluable for assessing the reach of TaRDIS outputs and informing dissemination strategies. Additionally, the Zenodo community strengthens TaRDIS's engagement with external stakeholders, including researchers, industry partners, and policymakers, by providing an organised and accessible collection of project results.

Have you explored appropriate arrangements with the identified repository?

At NOVA, GitLab is used as the primary repository for all TaRDIS source code and software artefacts, supporting version control and enabling collaborative contributions across teams. Hosting GitLab locally ensures data protection and reliability within NOVA's secure infrastructure. Additionally, NOVA developed a platform for storing and distributing datasets that also leverages open-source technology. This platform provides secure, long-term access to datasets and metadata via HTTPS, supporting both project needs and external access. By using open-source solutions, NOVA ensures flexibility, scalability, and alignment with open science principles, making TaRDIS data and tools accessible and reusable for the broader research community.

If there are restrictions on use, how will access be provided?

GitLab provides a comprehensive access control system that allows the flexible management of user permissions and restricts access to repositories using the git protocol. For datasets, the platform is accessible via HTTPS, and users may be required to register to track the impact of these datasets on the overall community.

Is there a need for a data access committee?

The DPO, NOVA, and the TaRDIS project coordinator assured that the DMP was implemented correctly and that no data access committee was required. We also note that the data used and produced by the TaRDIS project does not include user data, which mitigates ethical concerns about access to it.

Are there well-described conditions for access (i.e., a machine-readable licence)?

Yes, this was implemented. Additionally, all data elements made available by the project include open-source licences and indications of any restrictions on use and requirements for citing the TaRDIS European project (or specific outputs) whenever the data is used in research activities.

How will the identity of the person accessing the data be ascertained?

GitLab's access control mechanism ensures that only authorised individuals can access private repositories. Public repositories, by their nature, are accessible to everyone without authentication; however, as part of the security perimeter of the NOVA infrastructure, all access is logged by machines and stored for automated analysis to identify anomalies in access patterns.

Regarding the dataset repository, an online registry for accessing datasets was considered. This register serves only to collect data (with explicit user authorisation) on the number of accesses to each dataset and its intended use, for internal project monitoring. Similar to GitLab, as part of the security perimeter at NOVA, all access to all services is logged automatically and analysed periodically to identify anomalies in access.

2.3 MAKING DATA INTEROPERABLE

Are the data produced in the project interoperable, that is, allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e., adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

The data produced in the TaRDIS project is interoperable for use by the project members and for subsequent research after the project's termination. This has been achieved by consistently using open or widely used data formats for data exchange (with all formats clearly documented in the metadata). The concept of interoperable data demands that both data and metadata be machine-readable and that consistent terminology be used, principles that the consortium has adopted.

What data and metadata vocabularies, standards, or methodologies will you follow to make your data interoperable?

Data and research outputs have been described using standard descriptive metadata.

Will you use standard vocabularies for all data types in your dataset to enable interdisciplinary interoperability? In case it is unavoidable that you use uncommon or generate project-specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

Data and research outputs have been described using standard descriptive metadata.

2.4 REUSABLE INCREASE DATA RE-USE (THROUGH CLARIFYING LICENCES)

How will the data be licensed to permit the widest re-use possible?

The open data has been made available according to Open Licences such as Creative Commons.

When will the data be made available for reuse? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

The data is available for re-use upon the decision of the General Assembly, to avoid issues related to IP rights protection or access. The data remains accessible across the platforms described in this document for a minimum period that extends beyond five years after the conclusion of the project (although it is expected that the data will be available for a significantly longer period than this).

Are the data produced and/or used in the project usable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

One of the main results of the project is the TaRDIS toolbox, which is available under open-source licences and usable by the scientific community (researchers in academia and industry) in the field of Cloud-Edge Continuum, as well as software engineers and specialists in IoT, artificial intelligence, machine learning, distributed systems, and formal verification. The toolbox spans three major work packages: WP5 (distributed AI and orchestration tools), WP6 (the Babel runtime ecosystem, communication protocols, replicated data stores, and telemetry), and WP4 (formal analysis and verification tools for communication, data integrity, security, and federated learning orchestration).

Funded by the European Union

How long is it intended that the data remains re-usable?

All data remains publicly available for at least five years after the end of the project.

Are data quality assurance processes described?

Each partner bears the responsibility and must ensure data quality; raised questions were analysed by the steering committee.

Funded by the European Union

3 ALLOCATION OF RESOURCES

What are the costs of making data FAIR in your project?

How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions).

Who will be responsible for data management in your project? Who?

The project coordinator (NOVA) was responsible for overall data management, including maintaining the GitLab repository infrastructure, the dataset hosting platform, and the Zenodo community. Each work package leader was responsible for ensuring data quality and FAIR compliance for the outputs produced within their work package. The Data Protection Officer at NOVA (Dr João Marujo) oversaw GDPR compliance across all partners. Specifically, for the WP5 software tools, the tool developers (UNS, NOVA, NKUA, TID, and GMV) were responsible for maintaining the open-source repositories and associated documentation. For WP6, NOVA led the development and maintenance of the Babel ecosystem, communication and data management protocols, telemetry and monitoring tools, and experimental infrastructure; other contributors included UNS and TID. For WP4, the formal analysis tools were developed and maintained by UOXF (Scribble, session types), DTU (WorkflowEditor, Actyx middleware, compositional verification, PSPSP, CryptoChoreo, information flow), NOVA (VeriFx, Ant, AtomiS, JaTyC), and UNS (federated learning orchestration, JoinActors).

Are the resources for long-term preservation discussed (costs and potential value, who decides, and how what data will be kept and for how long)?

The project partners have allocated resources to cover the costs of open-access publications. Partners also used Open Access Publishing Platforms, namely arXiv and Zenodo, which incur no monetary cost.
The costs for data storage during and after the project finishes are the responsibility of the coordinator partner (NOVA). For NOVA, the costs are part of its research and development operation, and platforms referred to in this document are used by different projects at NOVA; the costs (hardware and human resources required to operate these platforms) are amortised across the multiple projects in which NOVA is involved. The open-source tools hosted on GitHub will remain available indefinitely at no cost. The Zenodo community deposits are maintained by CERN and are guaranteed to remain accessible for the lifetime of the repository. NOVA commits to maintaining the GitLab repositories and the dataset-hosting platform for at least 5 years after the end of the project, with the expectation that data will remain available for significantly longer. The General Assembly decided which datasets and outputs warrant long-term preservation based on their scientific value, reuse potential, and legal obligations.

Funded by the European Union

4 DATA SECURITY

Is the data safely stored in certified repositories for long-term preservation and curation?

Data is being stored in the previously identified repositories that ensure long-term preservation. As noted, the repositories hosted at NOVA are operated by NOVA's technical team, which ensures their continuous availability and data durability (through standard backup policies).

The security perimeter at NOVA includes several mechanisms that protect data from corruption or destruction by external malicious entities, as well as from unauthorised access. Most mechanisms within NOVA's security perimeter include Layer 2 packet inspection and automatic blocking. Furthermore, as previously reported, all platform accesses are automatically logged, and the logs are stored on separate machines. These logs are periodically inspected by machines to verify for anomalies, which can then be reported. Logs are stored for at least 1 year, enabling further analysis if an incident is reported. Backups are stored outside the primary infrastructure for additional security.

Funded by the European Union

5 ETHICAL ASPECTS

5.1 ETHICAL DIMENSION OF THE OBJECTIVES, METHODOLOGY, AND IMPACT

5.1.1 How TaRDIS addresses potential ethics risks

The TaRDIS consortium reviewed ethics issues as a standing item in the regular meetings of the project technical committee. The parties followed a structured approach to identifying, assessing, and disposing of ethical and data protection issues. All partners had equal responsibility for meeting ethical and legal requirements throughout their work on the project. The project had a dedicated task for ethics-related issues and monitoring, covering all phases. NOVA was responsible for managing the consortium's internal monitoring of the implementation of the ethics requirements.

Involvement of humans in research activities

During the pilot phase of the project, the consortium members tested the project outcomes. For this purpose, consortium members used anonymised or simulated data to assess the developed outcomes under operational conditions for the use cases. No human participants were recruited for this part of the project. Since all activities related to piloting involved only project partners, there was no need to recruit external participants or have them sign consent forms. During the three years of the project life cycle, whenever project partners deemed participation of individuals external to the project necessary, the consortium obtained (and clearly documented) their informed consent in advance. Participants were provided with Information Sheets and consent forms in a language and terms fully understandable to them, explicitly stating that participation was voluntary and that participants had the right to refuse to participate and to withdraw their participation or data at any time, without any consequences.

Personal data

The consortium ensured that, where the processing of personal data was involved, appropriate privacy-enhancing technologies were applied to the storage and transmission of that data between partners. Once the data had served its intended purposes within the project, it was deleted to mitigate the risk of accidental disclosure, unless it was required to be retained for legal or contractual purposes. All partners in the consortium adopted good-practice data security procedures in the project, helping to avoid unintended use or disclosure of data. Measures to protect data include access controls via secure authentication (including Layer 2 and application-level), installation of up-to-date security software on devices, regular data backups, as described in Section 2.
To comply with a potential request from an Ethics Panel or a Review, the TaRDIS coordinator has maintained the organisation's data protection officer for each partner in order to provide confirmation that all data collection and processing was carried out in accordance with EU and national legislation, if requested. Copies of this have been filed (by the project coordinator) and made available to the European Commission upon request.

5.1.2 Compliance with ethical principles and relevant legislation

All the data used in this project was pseudo-anonymised by the data providers. Artificial intelligence models were designed to be trustworthy and privacy-enhanced via federated learning strategies. Specifically, the federated learning approach adopted by TaRDIS (implemented through the Flower-based FL tool, PTB-FLA, and Fedra frameworks) ensures that raw data never leaves the local nodes; only model parameters or gradients are exchanged between participants. This privacy-by-design approach minimises the risk of data exposure. Furthermore, the FLaaS tool (T-WP5-10) was extended to support Differential Privacy (DP),

enabling a quantifiable trade-off between model accuracy and privacy guarantees. Split learning techniques were also implemented, allowing the distribution of model computation across IoT, edge, and cloud layers while keeping sensitive data at the edge. The lightweight ML tools (pruning, early exit, and knowledge distillation) further contributed to privacy by reducing the volume of information that needs to be transmitted across the network during training and inference.

Ethics

The TaRDIS consortium was fully aware of the ethical aspects of the research activities. It complied with the ethical principles and relevant national, EU, and international legislation. In particular, the consortium is committed to compliance with the Regulation Establishing Horizon 2020, Art. 19 "Ethical principles", Charter of Fundamental Rights of the European Union, the European Convention on Human Rights and the General Data Protection Regulation (Directive 2016/679) as well as principles ensuring the respect of human dignity, fair distribution of research benefits while protecting the values, rights and interests of the research participants. In this respect, in parallel with the development of the TaRDIS concept for preparing the proposal, the TaRDIS partners also emphasise identifying the ethical issues expected to arise.

For the personal data processed in the scope of the project, the applicable legislation is the GDPR. The partners gathered informed consent from any project participant, however in some instances, it was not necessary or possible to get informed consent given the following provisions in GDPR Art. 9.2: "*In case data processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject.*"
The acquired personal data was not, under any circumstances, used for commercial purposes or transferred to third parties or non-EU countries. In the event that such a need were to arise, transfers would only be made to jurisdictions that provide at least equal or higher levels of personal data protection. Had TaRDIS identified any data transfers lacking a lawful data transfer mechanism, partners would have reviewed the available options and promptly implemented the most suitable one.

Funded by the European Union