Project no.:	**101095738**		
Project full title:	**6G SHort range extreme communication IN Entities**		
Project Acronym:	**6G-SHINE**		
Project start date:	**01/03/2023**	**Duration**	**30 months**

D1.4 – DATA MANAGEMENT PLAN – THIRD VERSION

Due date	**31/08/2025 (M30)**	**Delivery date**	**29/08/2025 (M30)**
Work package	WP1		
Responsible author(s)	Berit Hvidberg Christensen (AAU)		
Contributor(s)	Enrico Maria Vitucci (CNIT) Baldomero Coll Perales (UMH) Gilberto Berardinelli (AAU) Saeed Hakimi (AAU) Ramoni Ojekunle Adeogun (AAU) Frank Burkhardt (FHG)		
Version	V1.0		
Reviewer(s)	Thomas Jacobsen (Nokia), Usman Virk (Keysight Technologies)		
Dissemination level	Public (PU)		

VERSION AND AMENDMENT HISTORY

Version	Date (DD/MM/YYYY)	Created/Amended by	Changes
0.1	03/07/2025	Berit H. Christensen	Transforming D1.3 into D1.4
0.2	06/07/2025	Berit H. Christensen	Updates based on COO input
0.3	07/07/2025	Berit H. Christensen	Updates based on contributors' input
0.4	15/08/2025	Berit H. Christensen	Version ready for internal review
0.5	19/08/2025	Berit H. Christensen	Updates based on reviewers' comments
0.6	22/08/2025	Berit H. Christensen	Minor refinements
0.7	26/08/2025	Berit H. Christensen	Proofreading and layout check completed
1.0	29/08/2025	Berit H. Christensen	Final version

Co-funded by the European Union

Horizon Europe Grant Agreement No. 101095738. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or SNS JU. Neither the European Union nor the granting authority can be held responsible for them.

TABLE OF CONTENTS

FIGURES

TABLES

ABBREVIATIONS AND ACRONYMS

.mat	MAT-files are binary MATLAB® files that store workspace variables
.py	Python Software
ADAS	Advanced Driver-Assistance System
CA	Consortium agreement
CARLA	Integrated High-Fidelity 3D Simulation Platform
CC-BY-NC	Creative Common Attribution-Non Commercial
CC 0	Creative Common Public Domain Dedication
CAVS	Connected and autonomous vehicles
DOI	Digital Object Identifier
DMP	Data Management Plan
EC	European Commission
EOSC	European Open Science Cloud
ETSI	European Telecommunications Standards Institute
EU	European Union
FAIR DATA	Making data findable, accessible, interoperable, and reusable
GA	Grant Agreement
GNSS	Global Navigation Satellite System
HTTPS	Hypertext Transfer Protocol Secure
IEEE	The Institute of Electrical and Electronics Engineers, Inc.
IPR	Intellectual property rights
ISBN	International Standard Book Number

ISRC	International Standard Recording Code
ITU	International Telecommunication Union
LOS/NLOS	Line of Sight/Non-Line of Sight
NEXT GA	Next Google Analytic
NGMN	The Next Generation Mobile Networks
O-RAN	Open (O) - Radio Access Network (RAN)
PoC	Proof of concept
SQL	Structures Query Language
SygMa	The EC Research Participant Portal
UTF-8	Unicode Transformation Format 8-bit
3GPP	The 3rd Generation Partnership Project
URL	Uniform Resource Locator
UTF-8	Unicode Transformation Format 8-bit
WP	Work Package
3GPP	The 3rd Generation Partnership Project (3GPP)
5G-ACIA	5G Alliance for Connected Industries and Automation.

EXECUTIVE SUMMARY

This deliverable, D1.4, represents the third and final version of the Data Management Plan (DMP) for the 6G-SHINE project. It builds upon the initial plan outlined in D1.2 and the subsequent updates provided in D1.3, incorporating insights and developments accumulated over the course of the project. While this deliverable is submitted at the end of the project (Month 30), it specifically addresses the data management activities carried out during the period from Month 13 (March 2024) to Month 30 (August 2025).

As the project advanced beyond its initial phases, the DMP was continuously adapted to reflect the evolving nature of data collection, processing, sharing, and preservation. D1.4 provides a retrospective and consolidated view of data management efforts during the second phase of the project, while also reflecting on how practices established in the first phases were sustained or refined through to project completion. It ensures alignment with FAIR principles, addresses compliance with ethical and legal requirements, and outlines measures taken to support long-term data accessibility and reusability.

The DMP for the 6G-SHINE project identifies and categorises the different types of data that are expected to be collected, processed, generated, and disseminated in the project, namely: public deliverables, scientific publications, contributions to standards, datasets, Software Code etc.

Furthermore, it describes the methods to ensure the generation of FAIR data, making it Findable, Accessible, Interoperable, and Reusable. The DMP describes the life cycle of the data and how it will be handled during and after the end of the project. In addition, it explains how the data will be selected for usage and presentation according to the project's data collection methods and standards, and finally, how the data will be organised and stored.

Compared to D1.3, this version includes a new Section 7.1, which addresses the Common Metadata Template for Datasets as required by the Smart Networks and Services Joint Undertaking (SNS-JU) for projects under its framework. The full metadata template is provided in Annex 1 for reference and implementation.

In this final version of the DMP, a summary of the data that has been collected and generated in the project, including the purpose of data collection/generation as well as the origin, type, and format of data, is presented.

1 INTRODUCTION

1.1 GENERAL

Efficient data management planning remains an aspect of research best practices, with the overarching goal of enhancing accessibility and promoting the reusability of research data. Complying with the Horizon Europe Model Grant Agreement, it is imperative to establish and regularly update a comprehensive Data Management Plan (DMP). This document is the deliverable D1.4 Data Management Plan (DMP), a product of WP1, Task 1.4 within the 6G-SHINE project. This DMP, presented in the 30 months since the project began, marks the final and third version on this topic.

The DMP for the 6G-SHINE project results from a collaborative effort involving the project coordinator and the valuable contributions and support from all project partners. It provides comprehensive insights into the strategies for data collection, processing, and generation within the various Work Packages (WP) throughout the project's 30-month duration. This DMP adheres to the template mandated by the European Commission for Horizon Europe projects [1]. In summary, the DMP outlines:

- The protocol for handling research data throughout and after the project's conclusion.
- The specifics of data collection, processing, and/or generation.
- The methodologies and standards to be applied.
- The intentions regarding data sharing and open access.
- The strategies for data curation and preservation, including post-project preservation.

As outlined in previous versions, the Data Management Plan (DMP) has been designed as a living document, evolving throughout the project's lifecycle to address emerging data types, methodologies, and evolving requirements. Over the course of the project, the DMP underwent two formal updates, D1.3 and this final version, D1.4, to ensure continued alignment with changes in policies, organisational structures, and other relevant factors affecting both the consortium and the broader research landscape. With the submission of D1.4 at the end of the project, the DMP is now finalised and will not be subject to further revision. Table 1 provides an overview of the revisions and updates made across the three versions of the Data Management Plan.

Table 1 Data Management Plan Overview

Deliverable No	Deliverable Name Work	Due Date (month)	Status
D1.2	Data Management Plan first version	M6	Submitted
D1.3	Data Management Plan second version	M12	Submitted
D1.4	Data Management Plan third version	M30	Submitted

1.2 The DMP Template

The European Commission has issued a template [1] to complete the DMP. The template contains questions to be answered according to the project. These questions have helped form the structure for the following chapters, where the template's main sections and proposed content are listed, including 6G-SHINE's responses.

1.3 Why is the DMP important?

The DMP is a crucial document within the project. It offers a comprehensive overview of the research data produced by the project, encompassing details about the data's types, formats, and the methods employed for processing and storage. This is done to ensure that the data adheres to the principles of FAIR data management, making it Findable, Accessible, Interoperable, and Reusable.

The primary purpose of the DMP is to provide guidance and assurance for appropriate data management practices throughout the entire duration of the project. Furthermore, the DMP encompasses:

- The consideration of any commercial and security aspects related to the data, identifying project data intended for sharing or open access, and information regarding the project's Intellectual Property Rights (IPR).
- The repository specification manages the data during and after the project's completion.
- A detailed description of the methodology employed in handling the data collected, processed, and/or generated within the project.

1.4 WHO USES THE DMP?

All project participants engaged in data collection, processing, and storage must respect the established data handling procedures and privacy notice. Additionally, all other project participants can refer to this document to gain insight into the data categories and their usage within the project.

1.5 FAIR PRINCIPLES AND OPEN RESEARCH DATA

6G-SHINE follows the FAIR guidelines for open research data to make its data Findable, Accessible, Interoperable and Reusable. These guidelines concern the data handling, evaluation, sharing and reuse of the manually and automatically generated input.

FAIR Data Management provides:

- Summary of the data collected.
- Methods to ensure that the data is FAIR.
- Resources to be allocated.
- Data security.

These principles precede the implementation choices and do not necessarily suggest any specific technology, standard, or solution.

2 DATA SUMMARY

2.1 IS EXISTING DATA BEING RE-USED?

Existing data refers to information from other projects and previous work carried out by the project partners, which may be reused with their discretion and permission. Some of the project partners (AAU, CNIT) have reused part of the simulation code developed earlier by themselves, and they have update it to generate results for the 6G-SHINE project. No datasets belonging to other projects or used previously by project partners have been reused in 6G-SHINE.

2.2 TYPES AND FORMATS OF DATA GENERATED

An overview of the data types collected and generated in the 6G-SHINE project is presented below:

- WP1/2/3/4/5/6: Publicly accessible project deliverables from different work packages
- WP2/3/4/5: Scientific publications (mainly scientific papers created by the project partners)
- WP2/3/5: datasets (training and testing algorithms)
- WP2/4: software code (open-source software, released in public repositories)
- WP2/3/4/5: contributions to standards
- WP1/2/3/4/5/6**:** other publications (promotion material, press releases, white papers, etc.)

According to the project's scope, the data types are labelled following the expected method of usage:

- O - Open; entries which are candidates for open research data.
- L – Limited; parts of data or post-processed versions shall be evaluated to see if they can be published as open research data, while the complete, unprocessed data are expected not to be made open.
- C – Closed; no part of this data is expected to be made accessible as open research data:

Datasets listed as 'Limited' or 'Closed' are restricted because making those openly available has been flagged as being against the beneficiaries' legitimate interests, as industrial partners expect commercial exploitation from these data.

2.2.1 Public Deliverables

All project deliverables are listed on our website (www.6gshine.eu) and made available for download following submission to the EC and approval by the project officer and external reviewers. Deliverables present the project advances in multiple technical fields and are provided in PDF format. Additionally, as requested by the chair of the SNS Steering board, the 6G-SHINE deliverables are also uploaded to the SNS repository on the BSCW server (Figure 3) upon submission to the portal. This upload includes the following disclaimer: "This deliverable has not yet been externally reviewed and approved by the European Union."

2.2.2 Scientific Publications

The scientific publications generated by the consortium partners are made available to the public and the scientific community in open source, using the 6G-SHINE Zenodo repository or other well-known open-science repositories such as arXiv (Table 4.) The publications are also listed in deliverable D6.3 "Final report of dissemination, exploitation and communication activities", and on the 6G-SHINE website.

2.2.3 Contributions to standards

6G-SHINE aims at pioneering the main technology components for short-range communications in in-X subnetworks. Given the project's timeline, which is expected to end earlier than the beginning of 6G standardization, the focus will be on pre-standardization. As described in 6G-SHINE's deliverable D6.1 [2], the project aims to contribute to the activities of major international standardization bodies and major industry fora. Potential contributions to pre-standardization are also presented in D6.1. Once a contribution to standardization was in preparation, this has been discussed within the consortium. In WP6, we have actively monitored and shared information on standardization activities. This tracking effort aimed to give all project partners a clear picture of the ongoing pre-standardization activities relevant to the topics of the 6G-SHINE project. So far, the 6G-SHINE project is tracking activities in 5 pre-standardization bodies and four standardization bodies, as listed in Table 2. Contributions to the

standardization bodies are presented in deliverable D6.3 “Final report of dissemination, exploitation and communication activities”.

Table 2 Pre-standardization and standardization bodies.

PRE-STANDARDIZATION BODIES		STANDARDIZATION BODIES		
1	5GAA	1	3GPP	RAN1 RAN2 RAN4 SA1
2	5G-ACIA	2	ETSI	ETSI DECT 2020NR ETSI ITS ETSI THz ISG ETSI RIS ISG
3	NEXTGA	3	IEEE	IEEE802.11 IEEE802.15
4	NGMN	4	ITU	
5	ORAN			

2.2.4 Other Publications

6G-SHINE aimed to reach a broad audience by generating different types of publications besides scientific publications. Those included flyers, posters, press releases, whitepapers, and information articles in national newspapers. These are also captured in deliverable D6.3 “Final report of dissemination, exploitation and communication activities “. There has been information on the website and/or 6G-SHINE social media channels about the release of these publications. Non-confidential publications have also been uploaded on the project website or linked to the websites where the publications are made available.

2.2.5 Datasets and Software code

The datasets and Software Code listed below have been made available for open access during the project.

Table 3 Summary of 6G-SHINE’s generated software and datasets.

WP (Partner)	**Name of Dataset**	**Short description and format**	**Expected file size**	**Label**
WP2/ Task 2.1/ Subtask 2.1b (UMH)	In-vehicle data traffic traces	Realistic data sets of raw and processed data traffic of in-vehicle networks that accurately represent the in-vehicle data traffic, especially the one related to the ADAS (Advanced Driver-Assistance System) domain of connected and autonomous vehicles (CAVs). The data sets differentiate the dataflow generated by different in-vehicle mounted sensors (cameras, radar, lidar, IMU, GNSS, etc.) and control functions by means of realistic traffic patterns (packet rate and size, traffic bursts, time correlation, etc.). The data sets are generated using an integrated high-fidelity 3D simulation platform	100 MB	O

WP (Partner)	Name of Dataset	Short description and format	Expected file size	Label
		(CARLA) with an automated driving stack (Autoware) available at UMH. (Software Code: Open-source software). Further details about the dataset are available at: 10.5281/zenodo.15655963		
WP2 Task 2.2. (CNIT)	MATLAB Simulation Package for Reconfigurable Intelligent Surface (RIS) Modelling Based on an Antenna Array-Like Approach	The dataset provides a MATLAB implementation of a macroscopic model for Reconfigurable Intelligent Surfaces (RIS), based on an antenna array-like approach. This model offers a simplified yet effective way to represent the electromagnetic behaviour of RIS without requiring full-wave simulations, which are often computationally expensive. The output of the model consists of electromagnetic field maps, enabling the visualization and analysis of how RIS alters the propagation environment. Matlab implementation of macroscopic model for Reconfigurable Intelligent Surfaces (RIS), which can be used to assess the realistic performance of RIS and integrated into ray tracing simulations, are made available in the dataset (Software Code: Open-source software).	Around 700 kB	O
WP2 Task 2.2. (CNIT)	Channel data for RIS assisted communication in a corridor scenario	The design of the 6G-SHINE radio system requires a deep understanding of the radio propagation characteristics in the scenarios where the subnetworks are anticipated to operate. The radio propagation characteristics could be affected by a high clutter density, obstructed line-of-sight. In such challenging environments, particularly in indoor industrial Reconfigurable Intelligent Surfaces (RIS) can be effectively deployed to enhance coverage and communication reliability. These surfaces act as passive or semi-passive reflectors that can be controlled to steer or focus radio signals on desired paths. The dataset consists of a set of propagation parameters obtained as output from ray tracing simulations of specific indoor scenarios. These parameters include angles of arrival, received power, and delay spreads, which characterize the radio propagation environment. In particular, the dataset focuses on indoor settings such as corridors, where signal reflections and obstructions are critical. Ray tracing simulations enable a detailed and accurate	Around 100 MB	O

WP (Partner)	Name of Dataset	Short description and format	Expected file size	Label
		comparison of the same scenario under different configurations, specifically with and without the presence of Reconfigurable Intelligent Surfaces (RIS). This allows for the assessment of the impact of RIS on the propagation characteristics, providing insights into how such surfaces can improve signal coverage and enhance received power.		
WP2 Task 2.2. (CNIT)	Scattering Measurements of Building Materials in the FR3 band	In both indoor and outdoor wireless communication scenarios, scattering is a key phenomenon that significantly influences how radio waves propagate. When electromagnetic waves encounter various surfaces, such building materials, they are partially scattered, altering the signal's direction, strength, and time of arrival. This results in complex multipath effects that impact fundamental propagation parameters like received power, delay spread, and angles of arrival and departure. The dataset includes detailed scattering measurements of various common building materials, collected in the 8–12 GHz frequency range (FR3 band). The materials tested comprise gypsum board, plasterboard with internal metal stud, plywood, brick wall, and marble. Measurements were conducted using a bistatic measurement setup, where the transmitter (Tx) was fixed at a 135° angle with respect to the normal of the sample centre, while the receiver (Rx) was rotated around the item under test (IUT) to capture the angular scattering response. The resulting data is provided in the form of Touchstone files, obtained through a Vector Network Analyzer (VNA), and contains the measured S-parameters for each material and configuration. These measurements serve as a reliable basis for modelling realistic electromagnetic interactions with construction materials in both indoor and outdoor environments.	Around 10 MB	O
WP4/ Task 4.1. (AAU)	Simulation code for radio resource management in hyperdense subnetwork scenarios.	The simulation code for mobile subnetwork scenarios is developed to model and perform the system-level evaluation of solutions in dense deployments of mobile subnetworks. The code base includes functionalities for creating deployments, simulating subnetwork mobility, modelling propagation, integrating algorithms for radio	Around 1 MB.	O

WP (Partner)	Name of Dataset	Short description and format	Expected file size	Label
		resource management and performing system-level analysis. (Software Code: Open-source software)		
WP4/ Task 4.1 (AAU)	Synthetic data traces for interference in dense mobile subnetwork scenarios	Simulated data sets of interference power in static and mobile subnetworks with different radio resource management techniques. The datasets capture the statistics of interference power in frequency, time, and spatial dimensions. The frequency dimension corresponds to different frequency sub-bands, and we capture the spatial dimension with the location of subnetworks inside simulated factory halls. The collected data sets contain time traces of aggregated interference power measured over each frequency sub-band at all subnetwork locations with the following dynamic sub-band allocations: random (fixed) sub-band selection, greedy channel selection, and centralised graph colouring. The data sets are generated using the in-X subnetworks simulator developed at AAU. The simulator implements functionality for system-level evaluation of mobile subnetworks and utilises channel models defined for indoor factory environments by 3GPP. (Software Code: Open-source software) By M12, a dataset for static dense subnetworks has been generated, considering the following setup: - 20 static subnetworks in a 20 x 20 m^2 areas, each serving a single device - Carrier frequency 6 GHz, 40 MHz bandwidth divided in *K*=4 channels - 3GPP InF channel model - 200.000 snapshots, with different subnetworks and device positions within the factory area. This preliminary dataset has a size of 2.38 GB. Further details will be included in an implementation note to be uploaded along with the dataset in the Zenodo repository.	Around 2.5 GB.	O
WP4/ Task 4.1 (AAU)	Simulation code and synthetic channel and mobility traces for dynamic subnetwork	This package includes a modular Python-based simulator and example datasets for dense mobile subnetwork deployments in industrial environments. Users can configure the simulator to generate custom datasets capturing subnetwork mobility, device-to-AP channel gains, and external interference across time and	Around 800 MB.	O

WP (Partner)	Name of Dataset	Short description and format	Expected file size	Label
	scenarios with optional external interference	frequency. Supported mobility models include parallel, random, and map-based movement (inspired by a 5G-ACIA factory layout). The propagation model follows 3GPP InF (Indoor Factory) channel specifications, with both dense and sparse clutter types, and incorporates path loss, correlated shadowing, and Gauss-Markov time-varying fading. External interference sources can be optionally enabled and can be either mobile or static; their temporal behaviour is modelled using either Poisson or on-off traffic patterns. Example configuration includes 10 mobile subnetworks in a 20×20 m² area, 3 frequency subchannels, and 10 GHz carrier frequency, but all parameters are fully configurable. The data is structured as NumPy arrays with snapshots over time, frequency subchannels, and subnetwork pairs, and is intended for benchmarking RRM algorithms, channel prediction models, and interference-aware scheduling strategies. All data is generated using the extended AAU in-X subnetworks simulator. Simulation code and datasets have been used for the ITU AI/ML challenge on 6G in-X subnetworks, launched by AAU in 2024. Further details are presented in deliverable D6.3.		

2.3 PURPOSE OF THE DATA COLLECTION/GENERATION AND RELATION TO THE OBJECTIVES OF THE PROJECT

The overall objective of the project is to pioneer the main technology components for in-X wireless subnetworks, short-range low-power radio cells to be installed in a wide set of vertical and consumer entities like robots, vehicles, production modules, and classrooms for the sake of supporting extreme communication requirements in terms of latency, reliability, or data rates. With a TRL 2-4, 6G-SHINE will leverage the opportunities offered by the peculiar deployment characteristics of such short-range subnetworks for a highly performant yet cost-efficient radio design that allows bringing wireless connectivity to a level of pervasiveness which has never been experienced earlier. The open-access datasets and code provided address relevant aspects of the project, such as traffic characterization for in-vehicle use cases, radio propagation in the in-X subnetwork scenarios of interest, and radio resource management. These datasets and software codes are described, as far as possible, in Table 3.

2.4 EXPECTED SIZE OF THE DATA

The size of the datasets made available in open access is also included in Table 3. The volume of all datasets was small enough to be stored on the standard available servers used in the project. There was no indication of a need to budget for repository storage outside of the scope of what repositories accept

within their free limit quotas. For one of the datasets, the WP4/Task 4.1 (AAU) 'Synthetic data traces for interference in dense mobile subnetwork scenarios', the dataset has a size of around 2.5 GB.

2.5 ORIGIN OF THE DATA

Data was generated by the 6G-SHINE partners' research activities. This data has been handled in accordance with agreements of usage (Consortium Agreement (CA) [3] in the project. For those partners providing open-access datasets, these are described, as far as possible, in Table 3.

2.6 DATA UTILITY

The data will be useful for subsequent research carried out by the partners, benefit the scientific community, and act as a basis for any subsequent projects.

3 FAIR DATA

3.1 MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA

3.1.1 Identification by a persistent identifier

All data deposited in 6G-SHINE Zenodo Community or other relevant repositories will be identified with a Digital Object Identifier (DOI) whenever possible to increase findability and citability.

3.1.2 Standards for Metadata

The project adhered to industry standards and will respect GA provisions regarding metadata: "Metadata of deposited data must be open under a Creative Common Public Domain Dedication (CC 0) or equivalent (to the extent legitimate interests or constraints are safeguarded), in line with the FAIR principles (in particular machine-actionable) and provide information at least about the following: datasets (description, date of deposit, author(s), venue and embargo); Horizon Europe funding; grant project name, acronym and number; licensing terms; persistent identifiers for the dataset, the authors involved in the action, and, if possible, for their organisations and the grant. Where applicable, the metadata must include persistent identifiers for related publications and other research outputs."

Each dataset has metadata associated with it, for descriptive, structural, and administrative issues. The embedded metadata in data are part of the data collection and documentation process, which are elaborated before the publication of the metadata, both in terms of sanitising and for adding metadata. Some metadata have been created during the publication procedure for a given repository and followed the (metadata) standard(s) for the specific repository. As for the specific metadata format, Zenodo utilises Data Cite [4], a de facto standard for describing datasets. All partners were requested to include the following acknowledgement of funding in their metadata, which is supported by Data Cite in the metadata standard in the grant section, linking to the grant number 101095738, or mention the name

of the project and grant agreement in case of metadata standards that do not have a separate grant section.

3.1.3 Search keywords

For generic repositories, the search keywords have been extracted from the metadata and provided as part of the descriptive metadata for the dataset. If subject-specific repositories are used, the search keywords followed the standard utilised by the given repository.

3.1.4 Metadata harvest and indexation

Metadata harvest and indexation will be available depending on the repositories' protocols. As for Zenodo, this is harvested into the most common search engines for research data like OpenAIRE [5].

3.2 MAKING DATA ACCESSIBLE

For publications, partners have been using The 6G-SHINE Zenodo repository (Figure 1) or other well established repositories such as ArXiv (Figure 2) or TechRxiv , as well as their own institutional repository. The 6G-SHINE Zenodo repository is also part of the EU Open Research Repository.

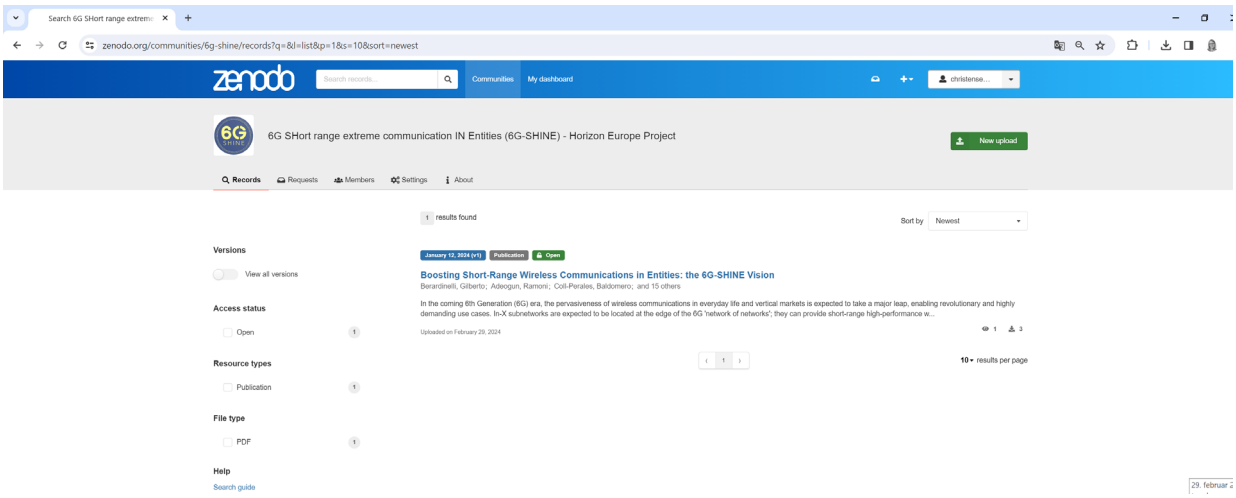


Figure 1 The 6G-SHINE Zenodo repository

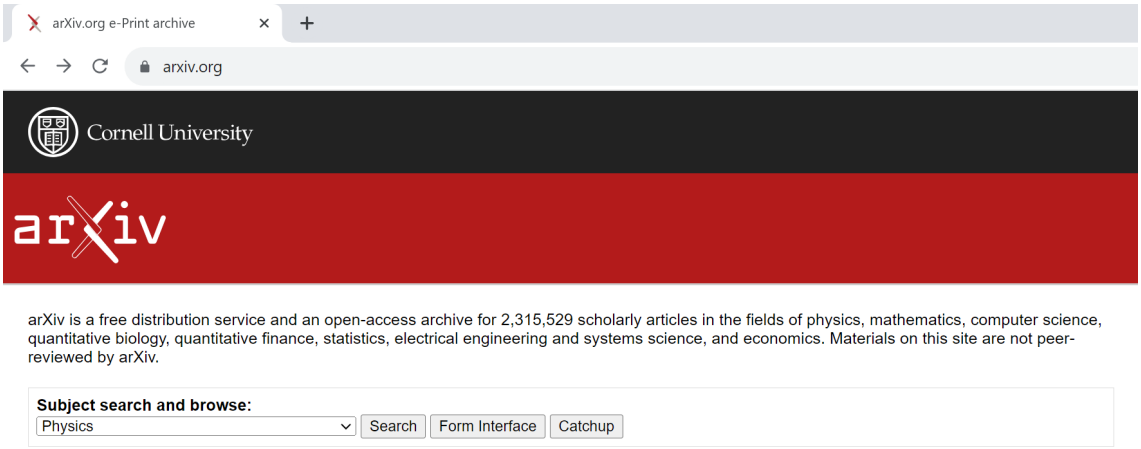


Figure 2 arXiv repository

A SNS repository has been setup on the BSCW server (Figure 3) by the SNS steering board, where all the project deliverables have been uploaded.

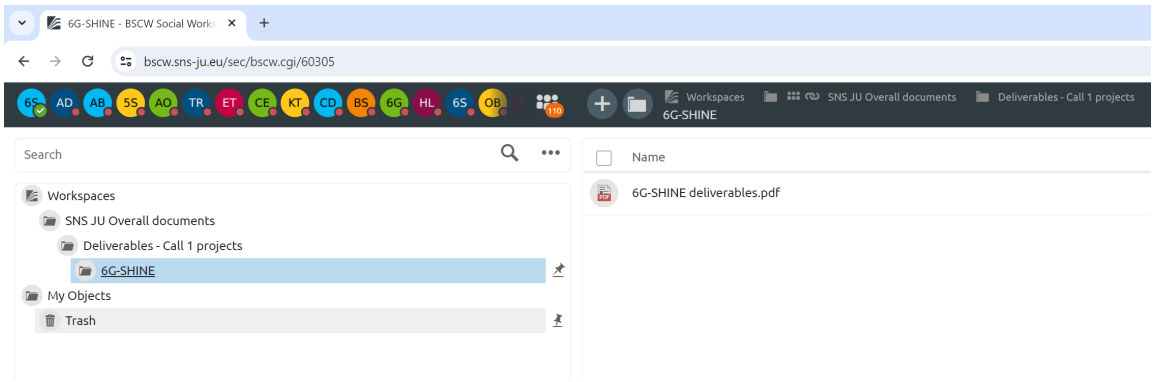


Figure 3 SNS repository on the BSCW server

Datasets deposited in Zenodo are provided with a DOI. Some of the institutional repositories that will be used, e.g., VBN [6], can provide a DOI upon demand. If this option is unavailable, partners have been instructed to deposit in Zenodo.

Repositories used by the partners are listed in Table 4.

Table 4 Repositories used in the project.

Partner	Repository
All	Zenodo is recommended by default for the 6G-SHINE Zenodo Community. For scientific publications, arXiv and TechRxiv can also be used.
All	The chair of the SNS Steering board stated during the February 12, 2024, meeting that all deliverables should be promptly uploaded to the SNS repository on the BSCW server upon submission to encourage collaboration across projects. The repository link is: https://bscw.sns-ju.eu/sec/bscw.cgi/14898

Partner	Repository
AAU	The general project information and relevant publications, are presented on VBN, Aalborg University's Research Portal. The code for the model description and some further Python scripts have been published to Zenodo to make the code citable with a DOI.
UMH	UMH, as a producer of the datasets, took responsibility for the storage and preservation of the traces. DOI was assigned, and data traces also made available on GitHub.
CNIT	CNIT, as a producer of the datasets, took responsibility for the storage and preservation of the traces. DOI have been automatically assigned to each data set via Zenodo or Figshare[7]. This ensures open access availability of all data for at least ten years and creates permalinks for all datasets, which can be used for referencing in publications.

3.2.1 Data

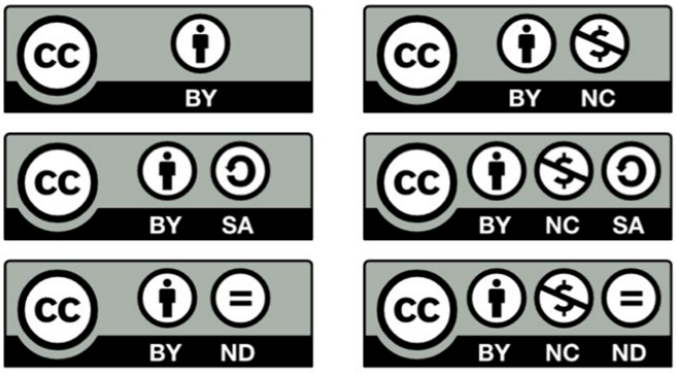
Data(sets) which have been made publicly available are identified in Table 3 (datasets marked as O – Open).

The data has been made available in various file formats, with some being open, whilst others with proprietary file formats due to the software used during the collection and analysis phase and to avoid the possible loss of important information if data is converted. The protocol for access is determined by the repository used.

For the project members, a common project platform is available in Teams, where the different datasets to be shared can be stored for internal use. Further, data only used by individual partners has been stored on their platforms and according to their internal rules.

Public data have been deposited in Zenodo or other relevant repositories depending on each partner's preferences. There has not been an active data curation procedure for the specific repository. Still, data should remain available and findable for five years after the end of the project (the Zenodo website mentions that the data will be available for at least 20 years). A need for a data access committee, e.g., to evaluate/approve access requests to personal and sensitive data, has not been identified.

For example, CC BY-NC has been used (see Figure 4). This license allows re-users to distribute, remix, adapt, and build upon the material in any medium or format for non-commercial purposes only, and only as long as attribution is given to the creator.



CC BY-NC includes the following elements:
BY Credit must be given to the creator
NC Only non-commercial uses of the work are permitted.
SA (Share Alike) Adaptations must be shared under the same terms.
ND (No Derivatives) No derivatives or adaptations of the work are permitted.

@ https://creativecommons.org/about/cclicenses/

Figure 4 The Creative Commons License Options

3.2.2 Metadata

Whenever possible, metadata have been made openly available and licenced under a CC 0, as per the Grant Agreement. The metadata includes information about software needed to access or read the data, and, when possible, a link to the relevant software has been provided.

3.3 MAKING DATA INTEROPERABLE

3.3.1 Version control

Each data owner was responsible for deciding on the file naming conventions and including a clear version-control guideline. Version control followed the file naming conventions that included a version control in the file name.

3.3.2 Standards to be followed.

All openly accessible data have been uploaded in a commonly accessible format and encoding (such as UTF-8). We are aware that some formats are considered proprietary. Where possible and feasible, we have converted files to non-proprietary formats. However, some data are best represented in the proprietary file formats, as some capabilities of the formats are not matched by their open alternatives.

3.3.3 Common vocabulary for data types

Data have been collected from many different partners; therefore, data types might differ depending on the provider. Therefore, the partners of the 6G-SHINE project intend to give the data relevant and representative names and use shared codes for representation in data, which clearly explained the content of the data type, including, e.g., the actual units in which they are measured, if relevant. This way, the partners can easily compare relevant data of the same type.

3.3.4 Qualified references to other data

Where possible, the documentation of the data included descriptions with references to other datasets. Furthermore, if needed, the metadata in Zenodo related to other datasets, e.g., pointing to related datasets, papers, etc.

3.4 INCREASE DATA RE-USE

3.4.1 Documentation is needed to validate data analysis and facilitate data reuse.

Partners were encouraged to provide documents such as e.g., readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, and units of measurement, whenever possible. The data and publications have been interlinked using metadata in Zenodo or other repositories where possible.

3.4.2 How the data will be licenced to permit the most comprehensive reuse possible.

For publicly published data, an appropriate license has been selected, i.e., CC-BY-NC license or equivalent. The exact license is to be decided by the data owner at upload time.

3.4.3 When the data will be made available for re-use?

Whenever it is possible to provide the data, it has been made available as soon as possible after its generation. The publication followed the publisher's guidelines if the data was supplementary to an article.

3.4.4 Usability by third parties of data produced in the project, after the end of the project

Parts of data have not been made publicly available and cannot be used by third parties. The data that was made publicly available can be used by other interested parties, with proper credit to creators and according to the specified license.

3.4.5 Documentation of data provenance

Throughout the 6G-SHINE project, data provenance has been documented in accordance with established good research practices across all participating disciplines. Each WP-team has followed discipline-specific protocols to ensure that the origin, processing steps, and transformations applied to datasets are traceable and transparent. These practices contribute directly to the reproducibility and reliability of the project's research outcomes.

Data provenance information is maintained through a combination of methods, including structured version control (e.g., GIT repositories for code and models), metadata annotation, and internal documentation protocols. For datasets shared publicly, provenance details are included in the associated metadata and documentation files. Where applicable, provenance is also supported by automated logs generated by software tools used in data processing and analysis.

These efforts ensure that the datasets produced within 6G-SHINE are not only FAIR but also verifiable and reproducible in accordance with open science principles.

3.4.6 Data quality assurance procedures

Data Quality Assurance requires a plan to reduce error and bias, data processing and validation errors, and to increase reliability and integrity. In this respect, data quality is acceptable if the collection methods are stable and protected from bias or manipulation, sufficiently documented and detailed, and cross-checked.

Quality control of the data in 6G-SHINE is the responsibility of the partner generating it, whilst quality assurance is a horizontal management activity of the project, led by the Project Coordinator. Details on 6G-SHINE's quality control procedures, which include internal review for deliverables' production, are set out in the Project Handbook [8], which partners should follow. No additional quality assurance measures at the partner's levels were required during the project lifetime.

4 ALLOCATION OF RESOURCES

4.1 COST ESTIMATION AND RESOURCES FOR MAKING THE DATA FAIR

Providing FAIR data is the responsibility of each partner, and a budget for open access fees has been included in the project budget for the partners who requested it. As for making data available through Zenodo, we remained within the free limits of Zenodo. We did not encounter extensive data handling

procedures for making data available, like converting data to new files, cleaning data or the like before publication.

4.2 DATA MANAGEMENT RESPONSIBILITIES

The project's General Assembly has handled potential issues regarding data management.
Project partners were responsible for the data generated; however, in case of disputes, the Project Coordinator was expected to be responsible for resolving these, aided by legal advisors. Nonetheless, this situation did not occur during the project lifetime.

4.3 COSTS AND POTENTIAL VALUE OF LONG-TERM DATA PRESERVATION

No additional cost was foreseen for long-term data preservation. All costs were included in the project budget.

5 DATA SECURITY

Each partner was required to follow the data security standards of their own institution, including guidelines for information security. This also included local guidelines for backup and restore procedures, as well as for ensuring, e.g., the proper user management for confidentiality, integrity, and accessibility of data. Where possible, partners used certified solutions for storing data. Partners transferred data using either secure encryption transportation protocols or by trusted encryption techniques, including proper transfer of encryption key(s), as necessary. All data storage was on the EU territory, except for the data generated by Interdigital, an associated partner residing in the United Kingdom.

6 ETHICAL ASPECTS

The project does not include any ethics deliverables. The usage of AI methods in the 6G-SHINE project does not bring any ethical issues in relation to objectives, methodology, and impact. The 6G-SHINE project will apply AI methods to data that is synthetically generated with simulations/emulation software platforms or measured in laboratory facilities. Such data refers to radio signal characteristics such as signal power, signal envelope, and data traffic measured in laboratory facilities related to control applications in machinery. For the latter, 6G-SHINE only measured and used data traffic characteristics related to packet size and packet interarrival times and not the content of the actual communication.

Since this project deals with the basic research design of radio system components, 6G-SHINE did not use any data related to humans and animals. The AI solutions developed in the project, therefore, do not stigmatise or discriminate against people. Also, the AI solutions are not expected to interact, replace or influence any human decision-making process since the designed solutions are only used for decision-making of radio electronics for the sake of improving wireless communication performance and reducing energy consumption. The AI solutions do not lead to any negative social and/or environmental impact nor raise any other ethical concerns.

7 OTHER ISSUES

Each partner is expected to follow their national and institutional procedures for data management.

7.1 COMMON METADATA TEMPLATE FOR DATASETS EMERGE FROM SNS-JU PROJECTS

In accordance with the decision made during the SNS Steering Board meeting held in Athens on 13 June 2025, all projects funded under the Smart Networks and Services Joint Undertaking (SNS JU) are required to comply with a common metadata framework for datasets generated within their respective scopes.
The 6G-SHINE project already ensures the storage and publication of its datasets through open-access repositories, including Zenodo, the 6G-SHINE project website, and the SNS BSCW server, in full alignment with the FAIR principles (Findable, Accessible, Interoperable, and Reusable). In addition to these efforts, it is now mandatory to submit metadata for each dataset to a centralized SNS JU metadata registry. This registry is intended to consolidate dataset information across all SNS JU-funded projects, thereby enhancing discoverability, interoperability, and ensuring the long-term accessibility of project results.
As communicated by the Technical Management and Validation (TMV) group, the metadata submission process is streamlined and user-friendly, estimated to require approximately 20 minutes per dataset. Metadata is to be submitted via an online form, with access credentials distributed to projects in early June 2025.
To support this process, the complete metadata template, outlining the required fields and structure, is provided in Annex 1 of this deliverable.
It is important to note that submission of metadata to the common registry is compulsory. Failure to comply might result in the relevant dataset not being officially recognized as a project result by the SNS JU. As of 9 July 2025, the SNS Online Metadata Registry became accessible to the Coordinators of SNS projects and was opened for dataset submissions. Accordingly, all open access datasets generated by the 6G-SHINE project had been submitted to the registry by the time of this report.

8 CONCLUSIONS

D1.4 provides a comprehensive overview of the data types generated and used by the project consortium, along with details on data formats, storage solutions, and how the data has been made FAIR (Findable, Accessible, Interoperable, and Reusable). The objective of this document is to report on the data management practices implemented throughout the 6G-SHINE project and to document the strategies used for data handling, reporting, and publication of project datasets.

Throughout the project lifecycle, daily data management has been a shared responsibility among all partners, guided by the principles and procedures described in this document. This collaborative effort involved close coordination between the project coordinator, the administrative project manager, the data-responsible partners, and the wider consortium to ensure consistent, compliant, and effective data practices.

As the final version of the Data Management Plan, submitted at Month 30, this deliverable reflects the culmination of all data-related activities within the project. It consolidates the lessons learned, tracks

the evolution of data management procedures across the project phases, and ensures that appropriate measures have been put in place to support the long-term accessibility and reusability of key datasets.

With the conclusion of the 6G-SHINE project, no further revisions of this DMP are foreseen. It serves as the final reference for all data management aspects undertaken during the project and supports the project's commitment to responsible research data stewardship beyond its duration.

ANNEX 1

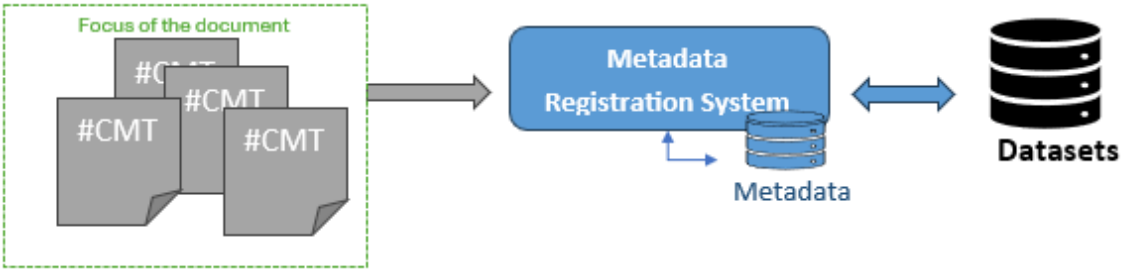
Common Metadata Template for Datasets emerge from SNS-JU projects

Common Metadata Template for Datasets emerge from SNS-JU projects

Introduction

The template includes a list of fields that refer to the **Metadata** of Datasets that emerge from the SNS projects. The fields should be completed in the process of storing a Metadata record of a Dataset in a **Metadata registry system**.

The Datasets themselves may be stored in any repository outside the Metadata registry system. Access to the repository where the Dataset is stored must be provided.



The proposed CMT structure is based mainly on a) what existing Metadata registry systems request during an addition of a new record (SLICES RI MRS, Zenodo, OpenAIREm, DataCite etc.), and b) what bibliography includes as vital information to best represent results coming from end-to-end experimentation processes (e.g., the ETSI TR 103 761 V1.1.1 and 3GPP TS 28.552)

In the proposed CMT, three main groups of information are requested, referring to:

- **Dataset Identity** (Basic fields regarding the ownership and the identity of the Dataset)
- **Dataset Object Characteristics** (Information on the Dataset model, size, availability etc.)
- **Dataset Content Information** (Dataset content description and its production process)

Introductory presentation: **https://bscw.sns-ju.eu/sec/bscw.cgi/606978**

1

Proposed Template

Dataset Identity fields

Field Identifier	Field description	Value type
Identifier	*Potentially produced by the system automatically*	
Version	*Potentially produced by the system automatically when an update to an entry is done*	
Name of related Dataset	e.g., "5G QoS Stress Test Campaign"	String/text
Date	Date when the data collected	Date
SNS Project Name	e.g., "FIDAL", "6G-SANDBOX" etc. (based on the implementation it can be a drop-down list for selection)	String/text
Owner Name	Name of Company or Consortium	text
Owner Contact email	e.g., contact@leader.com	email
List of Contributor(s)	e.g., Company A; Company B, etc (Semicolon separated)	List<string>
Dataset Short Description	Free text for the dataset owner to best describe the content of the dataset. E.g., 100words text. Public content listed in the MRS system (subject for search by the MRS)	Text
Dataset related publication	linked publications; e.g., https://doi.org/10.5281/zenodo.1234567 or IEEE explore link	List<string>
Dataset related Keywords	*Selection from a predefined list of Keywords maybe needed here to facilitate the search process*	List<string>

Dataset Object Characteristics

Field Identifier	Field description	Value type
GDPR and FAIR principles compliance statement	The dataset owner must confirm that the dataset is compliant with the FAIR and GDPR principles[1]	Boolean
License type	e.g., "CC-BY-4.0" reference licence types: https://creativecommons.org/share-your-work/cclicenses/ Related info: https://www.openaire.eu/how-do-i-license-my-research-data	String/text
External dataset link	This is the link to the repository where the actual dataset is hosted e.g., ["https://zenodo.org/record/123456"]	List <string>

Dataset Content Information

Field Identifier	Field description	Value type
A. Underlay Network & Compute infrastructure		
Underlay Platform	Reference link (if any) to existing platform used to collect the data (conduct the measurement campaign) e.g., link to Victoria Network UMA/ITIS	String
Environment	e.g., "indoors", "urban", "rural"	String
Network Infrastructure domain(s)	e.g., "RAN", "Transport", "CORE", "E2E"	String

[1] *A statement should be prepared and be available on the MRS system with a check box*

2

Compute Infrastructure type	e.g., "private-edge node" "central cloud"	String

RAN focused

RAN	RAN 3GPP Release	e.g., "Release 17"	String
RAN	New Radio Type	e.g., "NR-SA", "NR-NSA"	String
RAN	RAN split	e.g., "DU-RU split", "No-Split"	String
RAN	RAN focused technology	e.g., "O-RAN" "JSAC" "RIC" "No_focus"	String
RAN	Coverage Type	e.g., "Single_Macro", "Single_Micro", "Multicell_setup"	String
RAN	frequency band	e.g., "n78", "3.5GHz"	String
RAN	Bandwidth (MHz)	e.g., 20	Number
RAN	Max number of end devices involved	e.g., 5	Number
RAN	End devices mobility model	e.g., "static", "pedestrian", "vehicular"	String

Network Core focused

CORE	Network core Release release	e.g., "Release 17"	String
CORE	Network Core Solution	e.g., "OpenSource", "Commercial"	String
CORE			String

Transport network focused

Transport	Type of Transport	e.g., "wired", "microwave" "fiber_optics", "satellite"	String
Transport			

Compute environment focused

Compute	Orchestrator type	e.g., "Kubernetes", "OpenStack", "openSlice", "ONAP", "OSM"	String
Compute	gpu use	e.g., true	Boolean
Compute	Virtualization type	e.g., "KVM", "Docker"	String

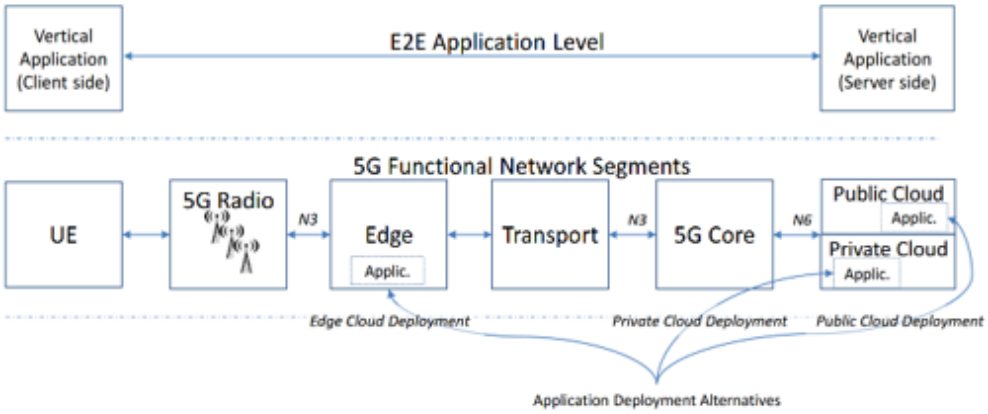
B. Service description		
Traffic origin	e.g., Manual network load (Iperf like), Application-originated traffic (e.g., video stream)	String
Traffic pattern	e.g., "UL/DL UDP", "UL/DL TCP", "UL/DL MPTCP", "DL UDP", "DL TCP", "DL MPTCP", "UL UDP", "UL TCP", "UL MPTCP",	String
Slice type	e.g., "slice101" "Multi-slice" "No_slicing"	String
Reference Plane	e.g., "control plane", "data plane", "management plane"	String
Related Vertical	e.g., "Vertical_agnostic", "CAM", "HEALTH"	String

C. Experimentation process and metrics		
Observation point (horizontal view)	e.g., select from the ones provided in Annex 1	String
Observation point (vertical view)	e.g., "Radio Level", "Network layer" "Application Layer", "Compute Resource-level", "cross-layer" "other"	String
Type of Measurement (Measurement Family	Select form those in Annex 2 or "other"	String
Measurement Tools used	e.g., tcpdump", "Prometheus exporter", ""	String
Measured indicators (add as many as needed)		
+Name of metric and units	e.g. Throughput (Kbps)	String
+Name of metric and units	e.g. delay (msec)	String
..	..	

Annex 1

Observation points for measurements (5GPP Arch WG View on 5G Architecture)

- E2E Application layer (client or server side)
- End device to Access code
- DU to CU domain
- Access node to edge node
- Access node to network core
- Network core to Cloud



Annex 2

The list of measurement families (3GPP 28.552):

- DRB (measurements related to Data Radio Bearer).
- RRC (measurements related to Radio Resource Control).
- UECNTX (measurements related to UE Context).
- RRU (measurements related to Radio Resource Utilization).
- RM (measurements related to Registration Management).
- SM (measurements related to Session Management).
- GTP (measurements related to GTP Management).
- IP (measurements related to IP Management).
- PA (measurements related to Policy Association).
- MM (measurements related to Mobility Management).
- VR (measurements related to Virtualized Resource).
- CARR (measurements related to Carrier).
- QF (measurements related to QoS Flow).
- AT (measurements related to Application Triggering).
- SMS (measurements related to Short Message Service).
- PEE (measurements related to Power, Energy and Environment).
- NFS (measurements related to NF service).
- PFD (measurements related to Packet Flow Description).
- RACH (measurements related to Random Access Channel).
- MR (measurements related to Measurement Report).
- L1M (measurements related to Layer 1 Measurement).
- NSS (measurements related to Network Slice Selection).
- PAG (measurements related to Paging).
- NIDD (measurements related to Non-IP Data Delivery).
- EPP (measurements related to external parameter provisioning).
- TI (measurements related to traffic influence).
- CE (measurements related to Connection Establishment).
- SPP (measurements related to Service Parameter Provisioning).
- BDTP (measurements related to Background Data Transfer Policy).

- DM (measurements related to Data Management).
- AFQ (measurements related to AF session with QoS).
- UCM (measurements related to UE radio Capability Management).
- PAU (measurements related to Policy Authorization).
- EEX (measurements related to Event Exposure).
- SDM (measurements related to subscriber data management).
- PPV (measurements related to parameter provisioning).
- DIS (measurements related to discovery).
- LM Location Management (measurements related to Location Management).
- SP (measurement related to service provisioning).
- DANS (measurements related to Data Analytics Service).
- OEU (measurements related to Network and Service Operations for Energy Utilities).
- UESA (measurements related to member UE selection assistance).
- AE (measurements related to analytics exposure).

REFERENCES

[1] Data management Plan template for Horizon Europe, https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/temp-form/report/data-management-plan_he_en.docx (EU Grants: Data management plan (HE): V1.1 – 01.04.2022).

[2] D6.1. Dissemination, Exploitation and Communication Plan for the 6G-SHINE project https://6gshine.eu

[3] General Contract Clauses: Consortium Agreement

[4] Date Cite https://datacite.org/

[5] OpenAIRE, a European project supporting Open Science https://www.openaire.eu/

[6] VBN, Aalborg University's research portal https://vbn.aau.dk/en/

[7] Figshare, online open access repository https://figshare.com/

[8] D1.1 Quality Project Management Handbook https://6gshine.eu