ENVELOPE

Evaluation and validation of connected mobility in real open systems beyond 5GS

Project deliverable D1.4

Data Management Plan

HORIZON JU Innovation Actions | 101139048 |
ENVELOPE - HORIZON-JU-SNS-2023

Co-funded by the European Union

6GSNS

Deliverable administrative information

Dissemination level	**PU - Public**
Type of deliverable	R – Document, report
Work package	WP1 – Project Management
Deliverable number	D1.4 – Data Management Plan
Status – version, date	V1, 21/06/2024
Deliverable leader	ISI-ATH
Contractual date of delivery	30/06/2024
Submission date	30/06/2024
Keywords	Storage; dataset; FAIR

Quality control

	Name	Organisation	Date
Peer review 1	Gabriele Scivoletto	NXW	28/06/2024
Peer review 2	Giulia Terracciano	COTO	28/06/2024

Version History

Version	Date	Author	Summary of changes
0.1	20-05-2024	Maria Trigka (ISI/ATH)	First draft of document structure and ToC, and the sections “Project Intro”, “Purpose of the deliverable” and “Intended audience”, and high-level descriptions for other sections
0.2	10-06-2024	All	Partners contribution in subsection 2.3.3
0.3	13-06-2024	Maria Trigka (ISI/ATH)	Updated project executive summary and the deliverable executive summary,
0.4	17-06-2024	Maria Trigka (ISI/ATH)	First complete draft shared for internal review

0.5	19-06-2024	Dimitris Fragkos (NCSRD), Pavlos Bassaras (ICCS)	Review and feedback that corrections are needed
0.9	21-06-2024	Lazaros Gkatzikis (ICCS)	Revision of document to include open call and network data aspects
1	30-06-2024	Lazaros Gkatzikis (ICCS)	Revision of document to address reviewer's comments

Legal Disclaimer

Table of contents

Deliverable administrative information I

Table of contents III

List of Tables IV

List of Figures IV

Project executive summary V

Deliverable executive summary 1

List of abbreviations and acronyms 2

1 Introduction 3

1.1 Purpose of the deliverable 3

1.2 Intended Audience 3

2 Data summary 4

2.1 Data categories 4

2.1.1 Technical data 4

2.1.2 Evaluation data 5

2.1.3 Open research data 5

2.1.4 Impact assessment data 6

2.1.5 Administrative data 6

2.1.6 Data on project outcomes and studies 6

2.2 Datasets description 6

2.2.1 Existing datasets 6
2.2.2 Data collection and generation in the context of ENVELOPE 1
3 Fair data 1
3.1 Making data findable, including provisions for metadata 1
3.2 Making data openly accessible 2
3.3 Making data interoperable 2
3.4 Increase data re-use 3
4 Allocation of resources 4
5 Data protection and ethical aspects 5
6 Data security 6
6.1 Secure data collection 6
6.2 Secure data storage 6
7 Conclusion 8
Disclaimer of warranties 1

LIST OF TABLES

Table 2.1 Existing datasets that could be used in ENVELOPE. 1
Table 2.2 Datasets to be generated by partners during ENVELOPE lifetime. 1

LIST OF FIGURES

Figure 6.1 ENVELOPE Data Repository in Microsoft Teams..... 7

Project executive summary

ENVELOPE aims to advance and open up the reference 5G advanced architecture, and transform it into a vertical-oriented one. It proposes a novel open and easy-to-use 5G-advanced architecture to enable a tighter integration of the network and the service information domains by

- exposing network capabilities to verticals,
- providing vertical-information to the network; and
- enabling verticals to dynamically request and modify key network aspects,

all performed in an open, transparent and easy-to-use, semi-automated way.

ENVELOPE will build APIs that act as an intermediate abstraction layer that translate the complicated 5GS interfaces and services into easy to consume services accessible by the vertical domain. The experimentation framework and the main innovations developed in the project are: MEC with service continuity support, zero-touch management, multi-connectivity and predictive QoS.

It will deliver 3 large scale Beyond 5G (B5G) trial sites in Italy, Netherlands and Greece supporting novel vertical services, with advanced exposure capabilities and new functionalities tailored to the services' needs. Although focused on the Connected and Automated Mobility (CAM) vertical, the developments resulting from the use cases (UC) will be reusable by any vertical. The ENVELOPE architecture will serve as an envelope that can cover, accommodate and support any type of vertical services. The applicability of ENVELOPE will be demonstrated and validated via the project CAM UCs and via several 3rd parties that will have the opportunity to conduct funded research and test their innovative solutions over ENVELOPE.

Social Media links:

@envelope-project

For further information please visit www.envelope-project.eu

Deliverable executive summary

This report describes the ENVELOPE Data Management Plan (DMP). The DMP contains a complete set of guidelines regarding the project data management with the primary objective of enabling free and online access to scientific information. It is a living document that will be updated at least 2 times along the project lifetime.

In particular, this report delivers an initial analysis of the data that will be generated and handled during the project lifecycle. It provides details about the storage location of the data, as well as qualitative characteristics of the datasets of interest. In addition, it defines the access policies for the different project outcomes (i.e., deliverables, publications and source code), and defines the data management processes.

The DMP is **organized** into sections covering as follows:

- Section 1 – Introduces the key concepts and purpose of the DMP.
- Section 2 – Presents the project's data description principles in a comprehensive manner to help partner beneficiaries provide their inputs with a high standard of quality.
- Section 3 – FAIR Data principles will be described to ensure that the project data is made findable, accessible, interoperable and reusable
- Section 4 – Presents the project's allocation of resources
- Section 5 – Analyse data protection and ethical aspects
- Section 6 – Discusses Data Security
- Section 7 – Conclusion that summarises the main points of the deliverable.

LIST OF ABBREVIATIONS AND ACRONYMS

Acronym	Meaning
B5G	Beyond 5G
CA	Consortium Agreement
CAM	Connected Automated Mobility
CM	Communication Manager
DL	Deliverable Leaders
EC	European Commission
GA	Grant Agreement
KPI	Key Performance Indicator
PU	Public
DMP	Data Management Plan
DMPO	Data Management and Protection Officer
FAIR	Findable, Accessible, Interoperable, and Reusable
GDPR	General Data Protection Regulation
IPR	Intellectual Property Rights
TMT	Technical Management Team
ToC	Table of Contents
UC	Use Cases
WP	Work Package

1 INTRODUCTION

1.1 Purpose of the deliverable

Deliverable 1.4 is the first version of the ENVELOPE Data Management Plan (DMP) that specifies how the data will be managed, securely stored and archived for further use by the research community. In the latter case, the DMP provides recommendations for future maintenance and access to the data by consortium partners and external parties. The DMP also specifies the procedures on how to handle personal data to guarantee citizens' fundamental rights and avoid misuse of the project results

Data are of fundamental importance in ENVELOPE and this DMP reports the types of data that are expected in the project providing some initial guidance about the data description and repositories, procedures and tools for data management, access and use in the ENVELOPE especially for research purposes. On top of data which will be generated by ENVELOPE itself as part of experimentation on the developed platforms by the consortium members or 3rd parties through the CAM UCs and open calls respectively, consortium members will provide existing data to be used in the project.

The DMP will ensure that the data is in line with the FAIR principles. Data privacy and security issues will also be addressed, where related requirements and applicable standards will be specified. The principle is to be as open as possible and to provide to the broader research community all data except those, which are conflicting with the law or commercial interests and IPRs of partners. The DMP will comply with EU and international regulations for the management and use of data and will be aligned with General Data Protection Regulation (GDPR). The main objective of DMP is to manage the research data in order to publish for research purposes (ORDP) a selected set of data, during the project and after the project completion. A Data Management and Protection Officer (DMPO) has been appointed to be part of the Technical Management Team (TMT). The document establishes procedures which are carried out through the following activities and for the entire duration of the project:

- Liaising with the TMT about the data management issues.
- Identifying and describing the data that will be collected, processed or generated.
- Defining what data will be made publicly available for research purposes and what sensitive data will not be shared because of Intellectual Property Rights (IPR) or confidentiality issues.
- Ensuring data protection and compliance with regulation on the protection of private data (GDPR).

This deliverable is the first version of the DMP and provides a high-level but comprehensive overview of the current knowledge fully achieving the ENVELOPE objectives related to this deliverable. An updated version of the DMP will be published in Deliverable D1.5 on month 24 of the project.

1.2 Intended Audience

The dissemination level of document D1.4 is 'public' (PU), thus the deliverable is available for all members of the consortium, the European Commission (EC) Services and those external to the project. This document is primarily intended to serve as an internal guideline and reference for all ENVELOPE beneficiaries, especially the governance bodies such as the General Assembly, the Technical Management Team, and the External Advisory Board. The main goal is for all beneficiaries to understand the procedures dealing with data management processes applied to the project.

2 DATA SUMMARY

This section provides an overview of the types of data that will be generated by the project. It is expected that ENVELOPE will also reuse existing data. This section contains an assessment of the ENVELOPE data at Month 06. It will be further detailed in the updated DMPs as the project developments progress. ENVELOPE will generate and manage various types of data, which can be categorized as follows:

- **Technical data**: Pertaining to the technical development and operation of the ENVELOPE reference architecture and the deployed UCs.
- **Evaluation data**: Involving testing and evaluation processes. This includes measurements of performance indicators and data required to validate the results presented in the project's outcomes and scientific publications.
- **Open research data**: Refers to scientific research results that has no restrictions on its access, enabling anyone to find, access, and re-use the corresponding data.
- **Internal administrative data**: Data generated and shared internally for administrative and management purposes.
- **Data on project outcomes and studies**: Data generated from managerial, technical, and scientific activities for reporting project achievements.
- **Impact assessment data:** Inputs from user questionnaire data for the evaluation of user acceptance of the ENVELOPE developments

A significant part of the ENVELOPE project data will be coming from third parties that will participate in the planned open calls. In particular, the projects that will be funded via Financial Support for Third Parties (FSTP) will also generate data. This data is categorized equally to the data of the project itself and will only be shared or made available according to the agreed levels of confidentiality.

2.1 Data categories

In the following paragraphs of this section, initial information is provided for the different areas of ENVELOPE research outputs and produced data (reflecting the status at M06).

2.1.1 Technical data

The technical data that is relevant for the DMP includes network and service data necessary for the project's development or operation, e.g., C-ITS data, sensor data, processed data, communication network data. It is anticipated that a great amount of data will be generated by the technical Work Packages of ENVELOPE. In particular, 3 experimental platforms will be used in ENVELOPE, one per site, offered by some of the project partners in order to generate important data for specific UCs to be exploited by the consortium. Data related to the description of UCs, requirements, performance metrics and a high-level ENVELOPE B5G platform reference architecture will be made openly available.

A large amount of data is expected to be generated by the systems used in the project (for example, by sensors and other systems from deployed automated vehicles or via network monitoring), but only part of this data will be used for actual development and validation. For instance, data generated by systems used in the project but not relevant to the research and development activities (such as vehicle data that do not pertain to positioning) will not be detailed in the DMP.

A significant amount of data is exchanged solely for operational purposes. Consider for example data from network monitoring which is a continuous process or each vehicle sending 10 cooperative awareness

messages per second, both resulting in large amounts of data. Storage of this data is not necessary for the project and would result in an unnecessarily large volume of data to be stored and managed. Consequently, only specific data identified as necessary will be made available, such as the data required for the verification of the developments. Concerning the latter, ENVELOPE will build on the experience gained in other EC-funded actions like the C-ROADS or the 5G-MOBIX tests, (where cooperative awareness messages have been exchanged as "pcap[1]" files). Alignment with the methodology applied by other SNS projects will be pursued and the previous related experience of many of the partners will be utilised.

2.1.2 Evaluation data

Evaluation data is a subset of the data produced during the lifecycle of the project and will be produced in the evaluation and demonstration tasks of WP6, including both the ENVELOPE and the open calls developments/UC. Such data will be produced from the internal tests conducted during the evaluation and validation of project UCs, and it will be accessible to stakeholders.

The data related to the evaluation of specific performance metrics, especially Key Performance Indicators (KPIs), is highly valuable for the entire community and is a crucial indication of the readiness of the implemented solutions to enter the market. The specific details of the evaluation data, such as its type and format, will be provided once the detailed definition of the evaluation procedures and KPIs is completed, and the specific data required for validating the project results is identified.

2.1.3 Open research data

Research data category contains all the data that result from research-oriented project activities, e.g., publications, validation data that can be used to reproduce the results. The principle is the data to be as open as possible and to provide all data except those which are conflicting with the law or commercial interests and IPRs of partners. The DMP will comply with EU and international regulations for the management and use of data and will be aligned with GDPR.

ENVELOPE consortium is committed to providing benefit to the European Community (EC) in terms of open access to scientific knowledge, standardization, and economic impact. Open access will be given to all scientific publications. All research publications produced in ENVELOPE will be deposited in repositories enlisted in OpenAIRE[2] (Open Access Infrastructure for Research in Europe), such as ZENODO[3], in a machine-readable electronic copy, at the latest upon publication.

In addition, the information about research outputs, tools or instruments required to validate the conclusions of scientific publications will be provided as well as the access to the results needed for validation. In the area ENVELOPE is addressing, cooperation is key. The general principle in ENVELOPE is therefore to build on open standards and share and discuss developments wherever and as early as possible. In order to make the results as relevant as possible, potential stakeholders are involved early, e.g., by input in standardization or by involving via liaison activities.

ENVELOPE will leverage the best practices implemented in other EC funded project, like the L3Pilot[4], on its data management and open data approaches designed to provide data to the Community, while protecting

[1] Data files mainly used in analysing the network characteristics of a certain data.
[2] https://www.openaire.eu/
[3] https://zenodo.org/
[4] https://l3pilot.eu/ European Project on piloting vehicle automation

IPR and all related exploitation aspects. Similarly, for the network open source platforms will be used and extended with specific ENVELOPE functionalities (e.g., Open5GS).

2.1.4 Impact assessment data

This category contains all the data related to the assessment of the impact of the project. The potential acceptance of the ENVELOPE platform and its UCs will be assessed via impact assessment/user acceptance questionnaire-based surveys. Questionnaires will be structured in the form of multiple-choice questions for facilitating easy processing and analysis. In limited cases and if the need is identified, a user may be able to enter free text, e.g., for providing more detailed qualitative feedback. In such cases, it will be carefully assessed whether those data and in which form can be made available due to GDPR. The details of the questionnaires and their exact format will be specified as part of Task 6.4 “Impact assessment and user acceptance”.

2.1.5 Administrative data

This type of data refers to information generated through project management activities, such as meeting minutes, recordings, internal reports, and data stored for historical purposes and follow-up. It is collected by the management team, which includes the project manager, the WP, and task leaders. Data is stored and shared between project partners using a project management tool that requires user authentication (ENVELOPE SharePoint). This includes i) internal administrative data and ii) FSTP administrative data.

Internal administrative data is accessible to all project partners, as described in the Project Management Plan (D1.1). Beyond this, the FAIR criteria do not apply to internal administrative data. The FSTP administrative data include all the data related to the overall management of open calls, such as documentation for inviting, evaluating and running the open calls. Notice that the results of the successfully concluded funded projects fall within the rest of the categories.

2.1.6 Data on project outcomes and studies

The data in this category is generated within the project to report project outcomes from managerial, technical, and scientific activities. This includes project deliverables, technical publications, videos, and presentations. This type of data may be useful for the general public, for example, for potential future projects that are built on the work of ENVELOPE, as well as for the industry and public bodies wanting to learn from ENVELOPE outcomes. All the data in this category is provided in text form and possibly in graphical representation. The data size is expected to be low.

2.2 Datasets description

In this section we present a first assessment of the relevant existing datasets and the new datasets to be generated by ENVELOPE.

2.2.1 Existing datasets

In the ENVELOPE project, publicly available datasets and inputs from standardization bodies will be used. The table below summarizes indicative useful existing datasets that can be used for generative modelling within the project. A non-exhaustive list of existing and/or publicly available datasets that could be reused by ENVELOPE is presented in Table 2.1. This table will be updated throughout the project's duration.

Table 2.1 Existing datasets that could be used in ENVELOPE.

Dataset owner	Dataset Name	Dataset description	Format/Type (e.g., .csv, json)	Size	Source link
Siemens, The Netherlands (SISW)	SynData	Synthetic data set, generated by the digital twin	.jpeg, .txt, json, .pcd	~ 1GB, compressed	Existing data set, used in the Envelope project by Siemens, internally stored, not publicly available.
AI4Mobile Project, Funding Agency: **German Federal Ministry of Education and Research**	Berlin V2X	GPS-located wireless measurements across diverse urban environments for both cellular (with two different operators) and sidelink radio access technologies, thus enabling a variety of different studies towards V2X.	.parquet	24.44 MB (cellular RAN)	https://ieee-dataport.org/open-access/berlin-v2x

2.2.2 Data collection and generation in the context of ENVELOPE

The ENVELOPE project will collect and generate data from diverse blocks within the B5G system and experimentation platforms supporting the UC trials, as developed by the three LLs. These data will help partners gain insights into the robustness of the methods and algorithms by evaluating well-defined indicators related to the KPIs of each UC. Data collection and generation will also support research and development activities in the project. In the next steps of the project, the developed methods and solutions in the innovation blocks (provided by the related work packages) will be further validated and improved with the data gathered from the project testbeds.

The collected/generated data constitute a valuable resource for studying, analysing and evaluating the performance of different infrastructure components, and developing new algorithms, models, and optimization techniques. With respect to the datasets that are expected to be generated by ENVELOPE's experimental platforms, Table 2.2 was filled by the involved beneficiaries. It includes the following information: dataset owner (namely, the partner who is responsible for the respective data), the related task and WP, the site (Greek, Dutch, Italian), the dataset name (at this stage of the project is optional) , a description of the dataset, its expected format, information of the open access storage medium (with a link, if available), if it is testbed applicable and if it is based on previous work.

Table 2.2 Datasets to be generated by partners during ENVELOPE lifetime.

Dataset owner (Site and partner name)	Relevant Task/WP	Dataset Name	Dataset description	Expected Format/Type (.csv, .json, xls, jpeg, pdf)	Open Access Data storage (Y/N)	Testbed applicable (Y/N), testbed name (if available)	Is this based on previous project /work (Y/N)
NCSRD	WP3/ T3.2	5GS location	Event Monitoring API notification, abstracted by ENVELOPE API, per cell id or geographical areas (e.g., zones)	JSON	Y	Greek Site	Y
NCSRD	WP3/ T3.1	5GC NF metrics	All the 5GC NF metrics supported by Open5GS solution (UPF, PCF, AMF, SMF)	Text/HTML, prometheus	N	Greek Site	N
NCSRD, OTE	WP3/ T3.1	-	KPIs collected during testing and demonstration of the Greek Site UC	text, Influx DB (line protocol), csv	Y	Greek Site	Y
NCSRD	WP3/T3.4	Compute resources	Compute resources collected from the edge/cloud infrastructure such as CPU, RAM, storage	prometheus	Y	Greek Site	N
ICCS	WP3/T3.2 WP4/T4.4	pQoS Datasets	Data sets with training data for the predictive QoS ML models. Features shall be extracted by the UE, the network e.g., gNb, the MEC servers. Data shall be recorded both on an infrastructure and a vertical service level.	.csv	Y	Y, Greek Testbed (ICCS/NCSRD)	N
eBOS	WP5		Open Calls Documentation material, results and beneficiaries' info	Pdf, word, excel, pptx, jpeg	N/A	N/A	N/A

SISW	WP3 and WP4	-	Data recorded during remote monitoring	.jpeg, .txt, .pcd	N	Dutch trial site	N
SISW	WP3 and WP4	-	Synthetic data generated out of the digital twin and recorded during mixed reality testing	.jpeg, .txt, .pcd	N	Dutch trial site	N
SISW	WP3 and WP4	-	Data generated out of the digital twin and recorded during teleoperated driving	.jpeg, .txt, .pcd	N	Dutch trial site	N
SISW	WP3 and WP4	-	Short video(s) for promotion of new technologies applied during testing and demonstration	.avi	N	Dutch trial site	N
SISW, TNO, KPN, CMS	WP3 and WP4	-	KPIs collected during testing and demonstration of the use-case	.txt, .csv	N	Dutch trial site	N
SISW	WP6, T6.4	-	White paper about Impact assessment and user acceptance results	.pdf	N	Dutch trial site	N
SISW, TNO, KPN, CMS	WP8	-	Project results and dissemination material	.pdf	N	Dutch trial site	N
TEO	WP4, T4.2	-	Raw cameras and Lidars data and data generated during road driving	,jpeg, .pcd, .csv	N	Italian trial site	N
LINKS, TEO	WP4		V2x messages generated during road tests		TBC	Italian Trial Site	N
HPE	WP4, T4.2, T4.4	5GC NF metrics	5GC NF metrics collectable from the dedicated API of the HPE 5GC	Prometheus	TBC	Y, Italian trial site	Y (API of the 5GC already available)

3 FAIR DATA

This section describes how data will be made findable, accessible, interoperable and re-useable (FAIR). The FAIR principles[5] are applied to research data and outputs in order to facilitate sharing and re-use of data, also in the long term, and to maximize the impact of research[6]. We delineate also how due to the specific characteristics of the data of each of the identified categories, the FAIR principles are applied slightly differently to each data category.

3.1 Making data findable, including provisions for metadata

In order to make the ENVELOPE research outputs and data findable, a set of measures will be carried out. First of all, to ensure findability the ENVELOPE data will be made available via direct links in the corresponding publications and deliverable. The links will be pointing to open platforms, like Zenodo, that support search and are tracked by the research community.

Second, metadata will be generated for the ENVELOPE data to facilitate automatic discovery of the public datasets that will be created by the project. Machine-readable metadata according to open-standard vocabularies will be provided along with those datasets. Metadata has significant impact on the findability of datasets. Notice that for published data sets, Zenodo stores metadata (and data) for the lifetime of the repository and provides the Open Archives Initiative Protocol for Metadata Harvesting API (OAI-PMH) this allows to harvest the metadata of the repository.

The project outputs will be made findable by using appropriate keywords together with the publications, and by linking them on the ENVELOPE website with respective keywords.

Summarizing, data will be made findable through:

- Direct links in research publications and deliverables.
- The ENVELOPE website: Data are categorized and tagged with keywords and links.
- Metadata and keywords at Zenodo or participant repositories.

Finally, all project findings will be presented to relevant stakeholder organizations. This will be done either via presentations or, where reasonable, via workshops. The potential extensions of any CAM or network features that require interoperability, in addition to the publication via deliverables and technical papers, will be input to standardization activities of 3GPP, ETSI and ISO. This increases the visibility of the ENVELOPE data and hence easier to be found as it ensures that the ENVELOPE contributions are part of an overall system/standard and not standalone contributions.

[5] Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, I. Jsbrand Jan; Appleton, Gabrielle; Axton, Myles; Baak, Arie et al. (2016): The FAIR Guiding Principles for scientific data management and stewardship. In: *Sci Data* 3 (1), S. 160018. DOI: 10.1038/sdata.2016.18

[6] https://www.openaire.eu/how-to-make-your-data-fair

3.2 Making data openly accessible

A fundamental component of the FAIR principle is to ensure that data is openly available to project partners but also to third parties. The publications-related datasets generated in ENVELOPE will be uploaded and stored in research data repositories; Zenodo and/or institutional repositories of project partners. Sensitive data, mainly related to IPR cannot be made available. The basic principle in ENVELOPE, as per the European Commission's guidance[7], is to make data 'as open as possible, as closed as necessary'.

Regarding technical and evaluation data, the following general procedure for defining data collection and storage has been established:

- Identification of the data relevant for logging, including the purpose and format of the data
- Definition of the logging scheme and format
- Definition of storage details and APIs

The identification of the data is currently being carried out and the definition of the logging scheme will follow. Discussions are underway to determine the exact data storage solution.

Research publications will be made openly available, given the publishing body copyright agreement, in a machine-readable electronic copy with open access, at the latest upon publication, and deposited in trusted repositories listed in OpenAIRE (Open Access Infrastructure for Research in Europe), namely Zenodo, and/or institutional repositories of project partners. Due to copyright issues some technical/scientific papers may instead be published via institutional repositories, which also follow the FAIR principles. Consortium partners will be free to choose between self-archiving ("green" Open Access) and open access publishing ("gold" Open access). Zenodo assigns a persistent identifier to publications made publicly available in the repository, and so do the institutional repositories of the partners.

Finally, user consent forms will not be made openly available as they contain sensitive personal data.

Beyond data, ENVELOPE promotes also the openness of its tools and research developments. Therefore, ENVELOPE adopts, as a general rule, to build on open standards and share and discuss developments wherever and as early as possible. Some project development will use open-source code and implement standard-based solutions or contribute to the open-source communities and standard specifications.

While the project will consider various options for making parts, if not all, of its results open access and available as open-source technology, the consortium is aware that challenges may arise that cannot be ignored and IPR protection will be seriously considered.

3.3 Making data interoperable

Interoperability is an important element of ENVELOPE as this is key to facilitate the interaction between the multitude of different involved entities, spanning both the network and services layers. On the network side, compliance with the messaging and data structures specified by 3GPP for 5G is fundamental. By utilizing and building on top of standards-compliant network deployments, either open-source like Open5GS or commercial ones, interoperability is guaranteed by design. ENVELOPE will target to contribute the necessary extensions to the most appropriate standard, which will serve as a certification of interoperability.

[7] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cuttingissues/open-access-dissemination_en.htm

The basis for the CAM research work at data level is the use of C-ITS messages; these are standardized in ISO and ETSI and the respective communication profiles are openly available and published at infrastructure and at vehicle level and will be used in ENVELOPE. This ensures interoperability and the correct interpretation of data. Overall, the developments of the project will build as much as possible on standardized elements.

Within the consortium the interoperability of published datasets and the respective metadata will be ensured through proper coordination at development and integration phases, and all ENVELOPE partners will be adopting common file formats, standards, vocabularies and methodologies. For technical data the most common exchange formats will be used such as JSON, XML, CSV etc.. Data processing and analysis will be mainly performed with open-source software.

3.4 Increase data re-use

It is the aim of ENVELOPE to make its research outputs (re)usable by 3rd parties and the overall research community. Our primary target groups are open call participants and subsequent projects in the Smart Networks and Services Joint Undertaking (SNS JU), or other research programs, will benefit from the availability of accessible datasets for within the scope of their activities.

To increase re-usability, each generated dataset will be formatted and will contain properly defined metadata to support easy machine-based analysis. Each dataset will be accompanied with a description file in human-readable format with a description of the dataset, detailing the type of data, its exact format, and indicative examples of use. Whenever compliant with a specific standard, the standard and its exact version will also be indicated.

The first round of open calls will help us evaluate whether the project data can be easily reused and if any issues are identified, corrective actions and measures will be taken, shaping the updated DMP by M24.

As ENVELOPE data will be made available via open platforms, they can be reused after the end of the project. This guarantees the long-term availability of the project data, and no need for additional plans for storage and curation has been identified.

4 ALLOCATION OF RESOURCES

At this stage, it is not possible to estimate the exact costs for making FAIR the generated and shared data and research outputs within ENVELOPE. The basic adopted principles are i) to use free software and platforms to store the collected data in the project, and ii) the partners to cover the expenses regarding managing repositories and uploading data from the already allocated budget.

The data used in ENVELOPE will be generated and collected as part of the project activities. Currently all project data are stored ENVELOPE SharePoint provided by ICCS, the Project Coordinator. Task 1.3 is dedicated to data management and is led by ISI/ATH, who also serves as the Data Management and Protection Officer (DMPO) of the project. The DMPO coordinates the data management activities, identifies potential issues, proposes solutions for ensuring compliance with data privacy and protection regulations, and collaborates with partners responsible for conducting the trials to establish procedures that guarantee the proper application of data protection policies at the national level. The DMPO is responsible for defining procedures to ensure that data is managed in accordance with FAIR principles, securely, and in compliance with GDPR.

ENVELOPE partners are accountable for collecting, managing, and sharing data in line with the guidelines and procedures outlined in the DMP. The budget for producing and publishing scientific and technical publications is allocated under Task 8.2 of the project. As for the long-term preservation of data, one cost-effective option is to store data in a repository such as Zenodo, which offers its services free of charge. The use of such open and free repositories will allow high visibility of the data and will ensure that project data will be available even after the end of the project.

Finally, legal costs are also expected to ensure that the collected data are compliant with local legislation, rules and GDPR.

5 DATA PROTECTION AND ETHICAL ASPECTS

This section outlines the approach to data protection and ethical considerations, and provides details on the methodology for achieving compliance with data protection regulations and ethical guidelines.

The project activities will comply with the Helsinki Declaration, the Ethics of Information and Communication Technologies report (2012) from the European Group on Ethics in Science and New Technologies to the EC, and the Charter of Fundamental Rights of the EU. All data processing activities will adhere to the requirements of the GDPR, as outlined in the Grant Agreement (chapter 15.2 - Data processing by the beneficiaries). Where it is possible without infringing legislation and directives regarding data protection and privacy we will provide - for the collected datasets - key feature extraction, storing, indexing and access through open interfaces. These datasets will be made available to the research community for future experimentation while taking into consideration privacy issues and proper anonymization of values following European guidelines such as those provided in the GDPR and by Article 29 of the Data Protection Working Party[8]. In addition, the information about research outputs, tools, or instruments required to validate and re-use research data will be provided.

ENVELOPE will conduct anonymous questionnaires among the participants of the UC demonstration. The questionnaires will not collect any personal data, only information about their opinions and attitudes towards the proposed demonstrated services and towards 5G and CAM in general. The user acceptance questionnaires will be collected locally by the local Living Lab partners, and only anonymized data will be used at the project level. The data processed will be relevant and limited to the project's purposes, following the "data minimization" principle. Based on ENVELOPE's objectives and methodology, there are no ethical concerns regarding the implementation and execution of the project, since no personal data is required in the user acceptance part, or in the evaluation and demonstration phase of the project.

Participants who are recruited for project studies will be able to give their informed consent and sign a consent form that is appropriate for their age and language before taking part in any new study or demonstration. The participants will not include any potentially vulnerable groups, minorities, or any other individuals who are unable to give informed consent. We will keep completed informed consent forms and information sheets, which explain the study in language that the participants can understand, on file. Participants will also have clear opportunities to provide feedback about their participation. The informed consent form will only contain the participant's name, surname, and potentially their age, and no other sensitive personal information.

Unavoidably certain personal data will need to be used for organizational purposes, namely for the realization of the open calls for FSTP. Such personal data will not be shared with unauthorized persons and will be handled in compliance with GDPR requirements. Therefore, appropriate security measures will be implemented to protect the integrity and confidentiality of shared data.

[8] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:HTML

6 DATA SECURITY

This section outlines how we plan to ensure data security. The ENVELOPE consortium is aware of the significant potential risks associated with data security, especially during data collection and exchange. These risks were identified and assessed at the beginning of the project in deliverable D1.2, "Quality Management Plan"

We will apply cyber-security and privacy principles to all data categories. This requires establishing general methodologies and principles for end users. Involving end users in test activities will be done responsibly by providing a briefing on the project and the test activities, including information about potential risks.

6.1 Secure data collection

The following security principles will be put in place to achieve the highest level of protection against any type of unauthorised modification:

- **Authentication:** Only authenticated users will be given access to data servers. Additionally, the servers themselves need to be authenticated.
- **Authorization:** Access to data servers will only be granted to authenticated and authorized users. User categories and rights will be predefined and strictly enforced. Access control mechanisms must be established for each trial site and project-wide to ensure proper authorization.
- **Accounting**: Every access and modification of a resource by any user needs to be securely logged to prevent user denial of access, alteration, or deletion of data files.
- **Confidentiality:** Data stored on servers need to be encrypted during transmission and storage to maintain confidentiality.
- **Communication Security**: All access to servers needs to be conducted through encrypted communication channels, such as Hypertext Transfer Protocol Secure (HTTPS).
- **Availability**: Security principles will ensure that server data is available to proper users during the designated service time. Regular data backups are essential.

6.2 Secure data storage

The official storing repository of the project (already fully operational) is the MS Teams/Sharepoint Online collaboration platform, hosted on Microsoft Azure cloud infrastructure. This shared space will serve as the reference data repository throughout the project's duration. The platform possesses all necessary features for secure data storage and access to the project data and fully complies with security and GDPR[9] requirements. Features as password protected login, groups, access and sharing rights, have been properly configured to ensure all security and GDPR requirements of the project. The deliverables and other relevant documents or data are uploaded to the appropriate Work-Package and Task subfolder under the Documents folder. Server databases and files are backed up daily according to Microsoft's policy.

The Microsoft Teams space is exclusively dedicated to ENVELOPE partners, and individuals outside the consortium cannot access it. The project manager oversees this tool and manages access rights to Teams.

[9] https://ec.europa.eu/easme/en/news/general-data-protection-regulation-gdpr-now-applicable-are-you-ready-it

Upon logging in with a username and password, users will be directed to the project's homepage via the following link (see Figure 6.1):

https://iccsgr.sharepoint.com/:f:/r/sites/Envelope/Shared%20Documents/General?csf=1&web=1&e=ZAFZux

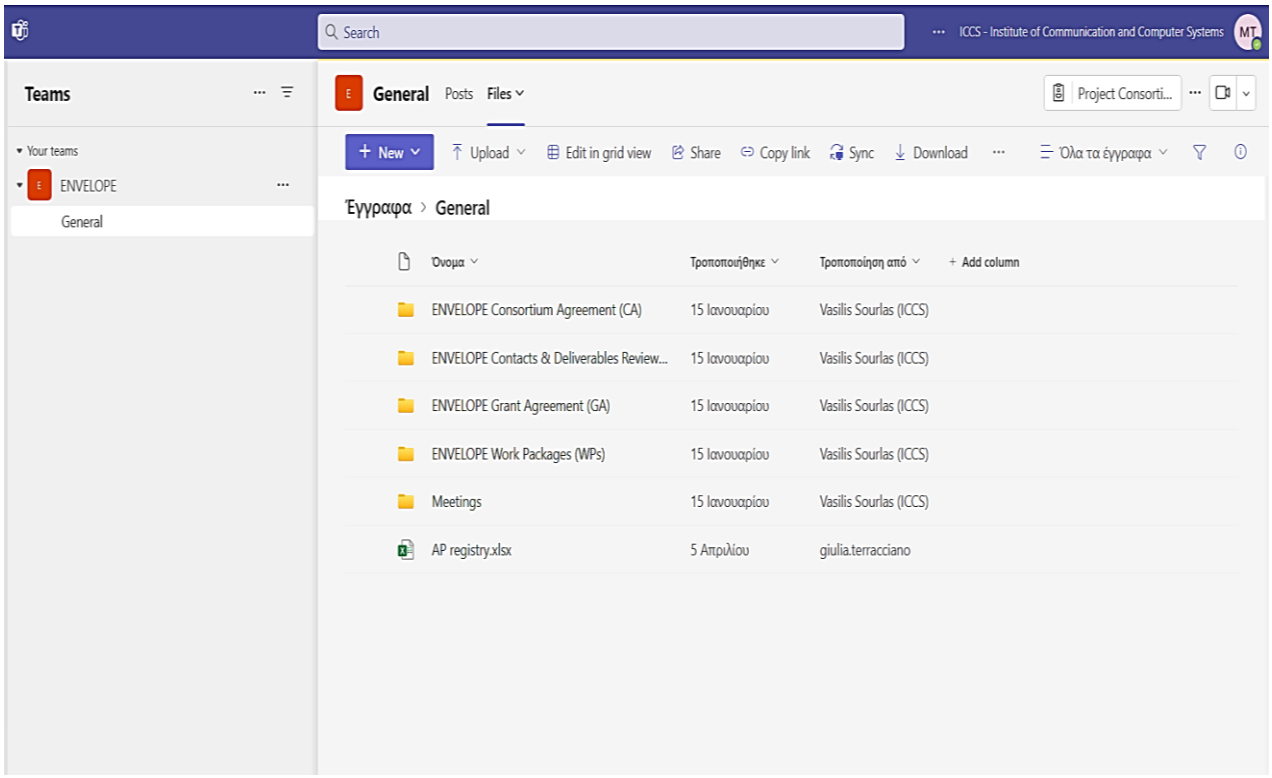

Figure 6.1 ENVELOPE Data Repository in Microsoft Teams

All data generated during the project (including those at the evaluation and demonstration and anonymised user questionnaire responses) will be also securely stored on the servers of each partner involved in data collection and generation. Third-party services like Dropbox and Google Drive will not be used for sensitive data storage and sharing.

Data produced during the ENVELOPE project will be stored in local and central servers throughout the project duration, treated as confidential, and subject to security processes and techniques to ensure compliance with GDPR regulations on data protection. Data will be shared among partners as determined appropriate by the partners themselves, and potentially sensitive data will only be stored locally in conformity with GDPR, the consortium agreement, and the internal consortium procedures as specified in D1.1 – Project Management Plan and D1.2 - Quality Management Plan.

7 CONCLUSION

This deliverable has described the initial version of the ENVELOPE DMP to be adopted in the project. It includes guidelines, an overview of the data that will be created or utilized inside ENVELOPE with high level description on the type and nature of the data managed by ENVELOPE partners. It establishes methodological principles that the project partners will follow, with a focus on making all types of data as FAIR as possible, while also addressing data legal and ethical considerations. This document is the first version of DMP and will be updated with additional information on the data management procedures. This document, together with the GA and the CA, is to be regarded as a reference for the ENVELOPE data management.

Disclaimer of warranties

The information and views set out in this deliverable are those of the authors and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the following information.

This project has adapted several methodologies already developed in previous projects, e.g., TARGET-X project (Grant Agreement number: 101096614) and PoDIUM (Grant Agreement number: 101069547). Ad hoc modifications were added to comply with the Grant Agreement conditions for ENVELOPE (Grant Agreement number: 101139048).