Fostering Opportunities Towards Slovak Excellence in Advanced Control for Smart Industries

D5.3. Data Management Plan

Date by 31.3.2023
v.1

Project: 101079342 — FrontSeat —HORIZON-WIDERA-2021-ACCESS-03

Funded by the European Union

TABLE OF CONTENTS

Funded by the European Union

DELIVERABLE INFORMATION

Work package	WP5
Task	Task T5.2: Data Management
Due date	31/03/2023
Submission date	28/03/2023
Deliverable lead	Ruhr University Bochum (RUB)
Version	v.1
Authors	Martin MÖNNIGMANN – RUB David MÜLLER - RUB
Reviewers	Gabriele PANNOCCHIA – UNIPI Miroslav FIKAR – STUBA
Dissemination Level	Public

Abstract	The project aims at increasing the research and academic prospects of Slovak University of Technology in Bratislava, Slovakia (STUBA) and at initiating the evolution of STUBA into a modern, reputed excellent institution that performs high-quality research in advanced automatic control, educates top-quality scholars and industrial practitioners, and is successful in active dissemination and exploitation of its research and innovation efforts. For this purpose, STUBA teams up with two renowned research groups in automatic control from RU Bochum, Germany (RUB) and Pisa University, Italy (UNIPI). The specific goals of the action are to reinforce the collaboration with the two research groups from Western Europe, to intensify research in advanced automatic control, to open up new collaboration channels through academic and industrial networking, to train excellent young/senior researchers and project managers, and to effectively disseminate and exploit the research results of STUBA. The unique features of the project are: - Adoption/amendment of internal research project-related rules and procedures and develop project management toolbox, - Research efforts aiming at the continued creation of high-quality research results and software tools, - Establishment of a series of guest scientific and academic lectures, - Exchanges and training of project managers and research (junior and senior) personnel, - Organisation of conferences and invited sessions, seminars with industry, and annual summer schools, - Preparation and implementation of a new PhD curriculum at STUBA, - Establishment of an academic-industrial research and innovation cluster.
Keywords	Control theory and optimization; Sensor networks, embedded systems, hardware platforms; Embedded systems; Monitoring and control systems;

Funded by the European Union

	Embedded systems in automation and control.

Document Revision History

Version	Date	Description of change	List of contributors

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

PARTNERS

The consortium of FrontSeat consists of 3 partners, as presented here below.

STUBA

Slovak University of Technology in Bratislava

RUB

Ruhr University Bochum

UNIPI

University of Pisa

Funded by the European Union

ABBREVIATIONS

Abbreviation	**Expanded Version**
DMP	Data Management Plan
FAIR principle	Findable, Accessible, Interoperable, Reusable
DEC	Dissemination, Exploitation, Communication

EXECUTIVE SUMMARY

The following document is Deliverable 5.3 of the FrontSeat Project, funded by the European Union`s Horizon Europe research and innovation programme under the action number 101079342. The purpose of the DMP is to support data management for all the data that will be collected, generated, processed and stored during the project in line with the FAIR principles. It provides information about the project´s data, storage, long time preservation and data policies. The document will be updated constantly until the final version in October 2025.

Funded by the European Union

1. DATA SUMMARY

Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.

If necessary, existing data will be used for scientific and industrial purposes available from previous joint projects of the partners.

What types and formats of data will the project generate or re-use?

The project will generate (I) Software training and testing data, which will use CSV for tabular data and XML, PDF/A, HTML, ASCII, UTF-8 for text data; (II) Laboratory experiments data and laboratory notebook, which will use CSV for tabular data, TIFF, JPEG 2000, PDF, PNG, GIF, BMP for still images and XML, PDF/A, HTML, ASCII, UTF-8 for text data; (III) Support data for publications (analyses and statistics), which will use XML, CSV for databases, MOV, MPEG, AVI for moving Images, ASCII, DTA, POR, SAS, SAV for statistics, TIFF, JPEG 2000, PDF, PNG, GIF, BMP for still images, CSV for tabular data and XML, PDF/A, HTML; ASCII, UTF-8 for tabular data; (IV) Support data for project activities (website hits, document downloads, survey results, analyses, statistics, etc.), which will use XML, CSV for databases, MOV, MPEG, AVI for moving images and TIFF, JPEG 2000, PDF, PNG, GIF, BMP for still images; (V) scientific publications, which will use PDF for texts; (VI) Training material, best practices, case studies, which will use DOCX, PPTX, PDF for text data, XLS for tabular data and MOV, MPEG, AVI for moving images; (VII) Materials produced for DEC purposes, DMP, which will use DOCX, PPTX, PDF for text data, XLS for tabular data and MOV,MPEG and AVI for moving images; (VIII) Contact lists (participants in summer schools, workshops, invited sessions, newsletter subscribers, etc.), which will use XML, CSV for databases. [1]

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

The data is primarily generated for research and testing purposes. Other purposes cover administrative tasks to implement the tasks, described in the Grant Agreement, properly. The data generation and collection will comply with the EU ethics and legal requirements as well as the national ethics and legal requirements.

What is the expected size of the data that you intend to generate or re-use?

The expected size of the data generated will be around (I) 10 MB for software training and testing data; (II) 50 MB for laboratory experiments data and laboratory notebook; (III) 500 MB for support data for publications, (IV) 500 MB for support data for project activities; (V) 50 MB for scientific publications; (VI) 50 MB for training material, best practices and case studies; (VII) 50 MB for materials produced for DEC purposes and DMP; (VIII) 5 MB for contact lists. [1]

What is the origin/provenance of the data, either generated or re-used?

Most of the data is primary data meaning it is generated by the researchers themselves or during experiments. Other possible origins are surveys, interviews, and subscriptions.

To whom might your data be useful ('data utility'), outside your project?

Funded by the European Union

The data generated during the project will be useful for other researchers/research communities, different companies, the European Commission and for education purposes.

2. FAIR DATA

2.1. Making data findable, including provisions for metadata

Will data be identified by a persistent identifier?

Data will be identified by a DOI. (Zenodo specific: e.g., 10.1000/zenodo.10000) [2,3]

Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Metadata of deposited publications must at least provide information about the publication (author(s), title, date of publication, publication venue), Horizon Europe or Euratom funding, grant project name, acronym and number, licensing terms. Metadata of deposited data must provide information at least about the datasets (description, date of deposit, author(s), venue, embargo), Horizon Europe or Euratom funding, grant project name, acronym and number, licensing terms. Metadata are structured as per the Metadata Standards Directory of the RDA group. [2,4]

Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

Keywords are properly set to optimize the possibility of discovery. [2]

Will metadata be offered in such a way that it can be harvested and indexed?

File and folders are named as the agreed conventions and file versioning is clearly understandable. [2]

2.2. Making data accessible

Repository:

The Repository used will be Zenodo. [2]

Will the data be deposited in a trusted repository?

Zenodo is a trusted repository, since it was developed by the Open Access Infrastructure for Research in Europe (OpenAIRE) and the European Organization for Nuclear Research (CERN) and it is currently funded by the European Commission. [2,3]

Have you explored appropriate arrangements with the identified repository where your data will be deposited?

No appropriate arrangements were necessary, because Zenodo was developed by an official European institution and is therefore fully European Open Science Cloud (EOSC) compliant.

Funded by the European Union

Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

The repository ensures that every digital output of the project will be identified by a DOI. All the DOI-identified materials will be reusable under a CC-BY licensing or a Creative Commons Public Domain Dedication CC 0. [2,3,4]

Data:

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.

(I) Software training and testing data can be accessed without restrictions; (II) Laboratory experiments data and laboratory notebook can be accessed without restrictions; (III) Support data for publications can be accessed without restrictions; (IV) Support data for project activities can be accessed partially restricted; (V) Scientific publications can be accessed without restrictions; (VI) Training material, best practices and case studies can be accessed without restrictions; (VII) Materials produced for DEC purposes and the DMP can be accessed without restrictions; (VIII) Contact lists can not be accessed.[1]
If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible. No embargo will be applied since the project does not plan to protect any intellectual property rights (IPR).

Will the data be accessible through a free and standardized access protocol?

The access to the data stored in Zenodo will be provided over standardized and free access protocols such as HTTP and OAIPMH. [3]

If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

There will be no restrictions on use during and after the end of the project.

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

It is not possible to identify the person accessing the data in Zenodo, since the repository generates an anonymized ID for every visitor changing every 24 hours, no matter if the visitor is signed in or not. Nevertheless, it will be differed between requests by humans, machines or robots, which are tracked separately.

Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

The General Assembly of the FrontSeat Project oversees evaluating/approving access requests to personal/sensitive data.

2.3. Making data interoperable

Funded by the European Union

What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

To make the data interoperable to allow data exchange widely readable data and metadata formats described in the data summary will be used.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

If it is unavoidable to use uncommon or generated project specific ontologies, mappings to more commonly used ontologies will be provided. The handling of generated ontologies or vocabularies will be discussed internally before further action will be taken.

Will your data include qualified references1 to other data (e.g. other data from your project, or datasets from previous research)?

The data generated during the project will include qualified references to other data if it enhances the overall quality of it.

2.4. Increase data re-use

How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

Information will be given via the repository about any research output or any other tools and instruments needed to validate the conclusions of the scientific publication. [4]

Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

Most of the generated data will be made freely available. The data will be licensed by the Creative Commons Attribution International Public License (CC BY) or the Creative Common Public Domain Dedication (CC 0), as per the Grant Agreement.

Will the data produced in the project be useable by third parties, in particular after the end of the project?

Produced data will be usable by third parties even after the end of the project under the accessibility policies outlined above, since Horizon Europe encourages the use of the R&I results through third party exploitation.

Will the provenance of the data be thoroughly documented using the appropriate standards?[1]

[1] 1 A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many

Funded by the European Union

The provenance of the data will be thoroughly documented using the appropriate standards (e.g. ReadMe Files).

Describe all relevant data quality assurance processes.

The quality of the data will be ensured via multiple internal reviewing processes.

Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.

3. OTHER RESEARCH OUTPUTS

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).

Every other digital output including software, workflows and models will be stored in the same repository, using the same standardized file types and policies, in line with the FAIR principle. In the event of generating new physical output, a proper documentation including all the necessary information will be setup. This documentation will be handled and published as a digital output using the same repository and policies in line with the FAIR principles.

Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

(See point above)

4. ALLOCATION OF RESOURCES

What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.) ?

Using the repository Zenodo does not cause any additional cost.

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

Costs will be covered as described in the Grant Agreement.

Who will be responsible for data management in your project?

meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source: https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/)

Funded by the European Union

The current Data Manager is David Müller from Ruhr-University Bochum (RUB).

How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

Long term preservation is ensured through the trusted repository Zenodo. Data will be stored for the lifetime of the repository, which will run for the next 20 years at least. Storing data in Zenodo is free of charge, regardless of the storage time. The person who is responsible for long time preservation will be announced during the project.

5. DATA SECURITY

What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?

All the data stored in the Zenodo repository will be secured by multiple backups on different disk servers in different locations and a system for automatic detection and recovery of file corruption on disks. Transfer of sensitive data can be realised using password protected files. Sensitive personal data should be anonymised with tools like OpenAIRE Amnesia. [3]

Will the data be safely stored in trusted repositories for long term preservation and curation?

Data will be safely stored for long time preservation due to the technical security system implemented at CERN (see point above).

6. ETHICS

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

The only ethical issue raised by FrontSeat, with respect to its research objectives, could be the collection of personal data. Personal data will be collected from the Newsletter distribution system, the system for public event invitations and from participants of the summer school. [5]

Intellectual property rights (IPR) may have an impact on data sharing, although the project does not plan to protect IPR according to Deliverable 5.3 (D5.3). [6]

Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?

In compliance with the GDPR, informed consent for data sharing and long term preservation will be included in questionnaires

Funded by the European Union

dealing with personal data, because, according to the GDPR, personal data may only be stored and processed as long as it is needed for the described purpose. [5]

7. OTHER ISSUES

Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

No other procedures for data management will be used.

Funded by the European Union