Project:

QUEST

(Grant Agreement number 101156088)

"Twinning to develop ELTE's research and innovation capacity in quantum reservoir computing"

Funding Scheme: Coordination and Support Action
Call: HORIZON-WIDERA-2023-ACCESS-02-02
Date of the latest version of ANNEX I: 24/05/2024

D8.1 FAIR Data Management Plan – initial version

Project Coordinator (PC):	Dr. Viktor Ivády Email: ivady.viktor@ttk.elte.hu
PC Organization Name:	ELTE
Lead Partner for Deliverable:	INT
Deliverable Due Date:	31/12/2024
Deliverable Issue Date:	18/11/2024

Document History

(Revisions – Amendments)

Version and date	Changes
1.0 – 18/11/2024	First version

Dissemination Level		
PU	Public	**X**
PP	Restricted to other program participants (including the EC Services)	
RE	Restricted to a group specified by the consortium (including the EC Services)	
CO	Confidential, only for members of the consortium (including the EC)	

QUEST's primary goal is to amplify the scientific prowess and innovative capabilities of Eötvös Loránd Tudományegyetem (ELTE) alongside its esteemed Twinning partners: Linköpings universitet, Loughborough University, Aalto University, JustInMind SL, and Intelligentsia Consultants Sarl. They collectively aim to create a quantum reservoir computing (QRC) system leveraging silicon carbide (SiC) defect qubits. Over three years, QUEST plans to roll out a research and innovation strategy centered on five core objectives, each associated with eight interrelated work packages.

- Objective 1 is geared towards pioneering research to construct a QRC system based on SiC defect qubits. This avantgarde technology could potentially enhance computational speed and diminish power usage by over 100 times compared to traditional machine learning models.
- Objective 2 emphasizes knowledge-sharing between ELTE's seasoned researchers and the Twinning partners, facilitated via short-term staff exchanges, seminars, and training sessions, all complementing the research from Objective 1.
- Objective 3 focuses on nurturing the career aspirations of early-stage researchers associated with both ELTE and the Twinning entities. They propose strategies like brief and intermediate exchanges, workshops, summer training sessions, and a collaborative PhD scheme.
- Objective 4 aims to refine ELTE's administrative and managerial proficiency for European R&D initiatives. Specifically, this entails enhancing the expertise of ELTE's Education Support Office and Research Support Office in aspects like project management, proposal drafting, intellectual property management, and scientific communication concerning European R&D funding programmes.
- Objective 5 is dedicated to bolstering the research stature and academic credibility of ELTE and its Twinning counterparts. This will be achieved through extensive communication, outreach, and dissemination efforts.

Table of Contents

1. Introduction

1.1 Scope of the deliverable

This document is comprised of the Data Management Plan (DMP) for the Horizon Europe funded QUEST project. The purpose of the document is to support the data management lifecycle for the data collected, processed and generated during the project, in particular with regard to the data needed to validate results presented in scientific publications.

The DMP describes how the research data is handled during the project, as well as will be handled upon the project completion. Namely, it explains what data is being collected (and its origin), generated and processed, if the data is publicly available or confidential, how the data is shared amongst the consortium partners and how it is curated and preserved.

The DMP will be updated regularly during the project (in any case, at least by M15 when deliverable D8.4 “FAIR data management plan – midterm version” will be issued, and M36, when deliverable D8.8 “FAIR data management plan – final version” will be issued). The updated versions will take into consideration the new data generated during the project’s activities. The document will also be updated in case of important changes to the project, such as the inclusion of new data formats, changes in consortium policies or other factors.

This document, which follows the template provided by the EC, is intended to be strictly compliant with the QUEST Consortium Agreement (CA) signed by all consortium members (ELTE, LIU, Lboro, JM, INT, and Aalto) and in particular with Sections 8 (Results), 9 (Access Rights) and 10 (Non-disclosure of information).

1.2 Structure of the document

The present DMP is structured according to the template provided by the EC and therefore includes the following sections:

1. Data Summary
2. FAIR Data
3. Allocation of resources
4. Data Security
5. Ethical Issues

2. Data Summary

Open research data applies primarily to the data needed to validate the results presented in scientific publications. In addition to this data, the present DMP considers also all the other data collected, generated, used or processed during the project, including e.g. promotional materials, presentations, minutes, reports, etc. As already stated in the introduction, all the data described in the DMP will be treated according to the policies defined in QUEST Consortium Agreement (CA).

Please note that the term “Confidential” in this document is not the same as “Confidential Information” in the QUEST CA. Data which is particularly sensitive and is considered to constitute Confidential Information (as defined in the QUEST CA) is not included in the DMP. The terms of exchange of such data between Consortium members shall be regulated by written agreements (e.g., NDAs) between Parties, according to Section 10 of the QUEST CA.

Also, the present DMP does not directly cover the implementation of open access to scientific publications produced in the framework of the project, considering open access to scientific publications is an obligation under the rules of Horizon Europe, to which every partner is expected to adhere.

The data collected, generated, used and processed during QUEST can be categorized as follows:

1. **Public Research Data**: any form of Non-Confidential data (possibly including data which has been previously produced and is not confidential) needed to validate the results presented in scientific publications and Non-Confidential products of research (including but not limited to designs, code, etc.) created and/or used in the framework of the project, where “Non-Confidential” means that such data can be made (or is already) publicly available.
2. **Confidential Research Data**: any form of Confidential data (possibly including data which has been previously produced) and products of research (including but not limited to datasets, designs, code, etc.) created and/or used in the framework of the project, where “Confidential” means it must not be made publicly available outside the Consortium.
3. **Project deliverables**: reports and draft reports.
4. **Promotional materials and templates**: presentation and report templates, posters, leaflets, press releases, website, etc.
5. **Official documents:** Proposal, Grant Agreement, Consortium Agreement, etc.
6. **Basic personal data** of project team members.
7. **Meeting data:** materials presented and/or created during consortium meetings, meeting minutes.

A more detailed description of the data is provided in Table 1 (Data Summary); this table will be constantly updated during the project, especially regarding Public Research Data, which will be subdivided according to format and purpose. Table 2 (Research Data description) provides some initial details on Public and Confidential Research Data.

Table 1. Data Summary

Data	Related WP(s)	Partners involved in generation/processing	Format	Expected size	Description
Public Research Data	WP1, WP2, WP3, WP4, WP5	ELTE, LIU, Lboro, JM, Aalto	.txt, .csv, HDF5, json, .npz, .pdf	< 1GB	Non-confidential data needed to validate the results presented in scientific publications and non-confidential products of research created/used in the framework of the project. It will mainly consist of datasets from experiments / simulations and code.
Confidential Research Data	WP1, WP2, WP3, WP4, WP5	ELTE, LIU, Lboro, JM, Aalto	.txt, .csv, HDF5, json, .npz, .pdf	< 1GB	Form of data and products of research (including but not limited to designs, datasets, code, etc.) created/used in the framework of the project and considered confidential according to QUEST IPR protection policy.
Project deliverables	All WPs	All Partners	.doc, .docx, .pdf	< 1GB	Confidential and public deliverables listed in the Grant Agreement
Promotional materials and templates	WP8	INT	.psd (Photoshop), .indd (Indesign), .jpg, .png, .pdf, .ppt/pptx, .doc/docx, .mp4 and similar formats	< 10GB	Website, leaflets, posters, logos, PowerPoint presentations, films, templates for presentations and deliverables, etc.
Official documents	WP9	All Partners	.pdf	< 1GB	Grant Agreement (GA), Consortium Agreement (CA), Project Proposal and all official documents related to the project.
Basic personal data	WP9	All Partners	.xlsx	< 10MB	Excel file containing basic personal data (name, surname, organisation, phone and mobile numbers, email, skypeID) of the staff involved in the project
Meeting data	All WPs	All Partners	.pptx, .pdf, .doc/docx and similar formats	< 2GB	Presentations, minutes and any other material discussed or created during project meetings

Table 2. Research Data Description

Partner Short Name	WP#	Data description				Tools for accessing and/or processing	Data utility	Confidentiality (PU/CO)
		Purpose	Origin	Format(s)	Size			
ELTE	1-5	Simulation data	Running simulations of point QRC model	.txt	Several GB	Text editor	All WP1 and WP3 groups	CO (unless published)
LIU	1-5	Simulation models: System parameters and simulation inputs	Own model development work and customized scripts	.txt, Python- or Matlab-readable data-array files (npz)	A few GB	Text editor or provided readout Python scripts	All WP1/3/4 groups	CO (unless published)
LIU	1-5	Simulation benchmarking	Data is generated using customized scripts	.txt	Several MB	Text editor	Researchers	PU
LIU	1-5	Simulation data	Running simulation software with developed models	.txt, Python- or Matlab-readable data-array files (npz)	1-100 GB / model	Text editor or provided readout Python scripts	All WP1/3 groups and paper writing	CO (unless published)
LIU	1-5	Experimental schematics	Own experimental setup documentation	.txt, images (.pdf, .jpg, etc.)	10-100 MB / setup	Text editor or image viewer	Internal use / paper writing	CO (unless published)
LIU	1-5	Experimental data	Running experiments on sample	.txt, Python- or Matlab-readable data files, or images (.pdf, .jpg, etc.)	A few GB per experiment	Text editor, provided processing scripts or image viewer	All WP3/4/5 groups / paper writing	CO (unless published)
LIU	1-5	Sample/Device characterization data	Own optical characterization	.txt, binary	1-100 kB	Text editor, home-developed graphic programs	WP3/4/5	CO (unless published)
Lboro	1-5	Simulation models and data	Own model development work and customized scripts	.txt, Python- or Matlab-readable data-array files (npz)	1-100 GB / model	Text editor or provided readout Python scripts	All WP1/3 groups and paper writing	CO (unless published)
JM	1-5	QRC model parameters. Parameter tuning for QRC models	Model development work	JSON	100-500 kB	Python, Quantum Computing SDKs	Researchers, Industry and applications	CO (unless published)
JM	1-5	QRC training data. Training models	Computational experiments	.csv	10-50 GB	Python, Quantum Computing SDKs	Researchers, Industry and applications	CO (unless published)
JM	1-5	QRC benchmark results. Comparing QRC performance with classical	Benchmarking experiments	.csv	1-5 GB	Python, Quantum Computing SDKs	Researchers, Industry and applications	CO (unless published)
JM	1-5	Enhanced deep learning Models. Optimized models using QRC	Model Optimization	HDF5 or similar format	100-500 MB	Python, Deep Learning Frameworks	Researchers, Industry and applications	CO (unless published)
JM	1-5	QRC Hardware performance logs. Monitoring and debugging QRC hardware.	Hardware testing	.txt, .csv	500 MB-2 GB	Python, Logging Tools	Researchers, Industry and applications	CO (unless published)
Aalto	1-5	Simulation data	Own model development	.txt	Several MB	Matlab, Mathematica	All WP1 groups	CO (unless published)

3. FAIR Data

3.1 Making data findable

Data will be made findable (through DOI automatically assigned by Loughborough University's research data repository[1] and Zenodo[2]) if it is used for validation and support of scientific publications; for this data, metadata will be created according to standards before uploading on the Loughborough University's research data repository or Zenodo. Description of such metadata will be provided in future versions of the DMP and will likely follow standards (e.g. DCMI[3]). Public deliverables and promotional materials will be made available through the project website (quest.elte.hu). Confidential data will not be made findable and will be provided with metadata only if required by partners.

Data	Naming Conventions and Folder Structure	Metadata
Public Research Data	To be defined in future versions of the DMP	Data sets will be documented with appropriate metadata once stored on 4TU.ResearchData repository
Confidential Research Data	To be defined in future versions of the DMP	No metadata is expected to be used for confidential research data. Eventual metadata files will be added by the owner of the data if deemed necessary.
Project deliverables	Final versions of Deliverable reports will use the following naming convention: D*X.Y_DeliverableTitle*.pdf; draft reports will use the same naming conventions but will be in docx format and will contain the date of last save at the end of the file (D*X.Y_DeliverableTitle_YYMMDD*). Deliverable reports and draft reports will be shared on Loughborough University's servers and grouped in folders according to WP. Public deliverables will be available for download from quest.elte.hu	No metadata is expected to be used for project deliverables.
Promotional materials and templates	Final versions of promotional material will use the following naming convention: QUEST_*Description.ext*, where *.ext* is a generic extension. Final versions of templates will use the following naming convention: DX.Y_Deliverable_Template.docx, Presentation_WP#_Template.pptx. Final versions of promotional materials and templates will be shared on Loughborough University's servers and grouped in 2 specific folders: "PromotionalMaterials" and "Templates", which might be further subdivided into subfolders. Also, some promotional materials will be available for download from quest.elte.hu	No metadata is expected to be used for promotional materials and templates
Official documents	No naming conventions are expected to be used. Official documents will be shared on Loughborough University servers and grouped in a specific folder: "OfficialDocuments"	No metadata is expected to be used for official documents
Basic personal data	Basic personal data will be collected in single Excel file: QUEST_TeamMembers.xlsx; it will be shared on Loughborough University servers.	No metadata is expected to be used for basic personal data.
Meeting data	The following naming convention will be used: (i) Presentations: *YYMMDD_WP#_partnershortname*.ppt/pptx/pdf; (ii) Minutes: *YYMMDD_M#*_Minutes.pdf/doc/docx; (iii) Other files: *YYMMDD_Description.ext*; (iv) Meeting Folders: *YYMMDD_MeetingVenue*; where *YYMMDD* is the date of the meeting and *.ext* a generic extension. Materials presented during consortium meetings and meeting minutes will be shared on Loughborough University's servers and grouped in a specific "Meetings" folder which will be further subdivided into folders related to the specific meeting	No metadata is expected to be used for meeting data.

Table 3. Naming convention and metadata

[1] https://repository.lboro.ac.uk
[2] https://zenodo.org
[3] http://dublincore.org/documents/dces

Table 3 illustrates the approach to making public data findable for the different sets of data described in Section 2 ("Data Summary"). Also, it describes the naming conventions and folder structure used for files present on Loughborough University's servers, where a storage space will be created and used throughout the project for data exchange. Part of the data uploaded on Loughborough University's servers will be confidential and access to this data will be regulated with different authorisation levels (see Section 3.2, "Making data openly accessible"). Version numbers will be used only for draft reports and will coincide with the date of last save. Keywords for publicly available data will be defined in future versions of the DMP.

3.2 Making data openly accessible

As a default, the only data made publicly available will be: (i) Non-Confidential data needed to validate the results presented in scientific publications and Non-Confidential products of research created in the framework of the project; (ii) promotional materials; and (iii) public deliverables. The data in (i) will be made available through storage in Loughborough University's research data repository or Zenodo, while those in (ii) and (iii) will be shared among partners on a secure server maintained by Loughborough University and/or made available for download from the project website (quest.elte.hu). All the other data generated, collected, used or processed in the framework of project activities will be considered Confidential as a precaution. Confidential data will be shared between partners using a dedicated storage space on Loughborough University's servers; this space will be accessible through usernames and passwords communicated to each team member. There will be 3 levels of authorisation: (i) Steering Committee (full access, for members of the Steering Committee and EC); (ii) Project Technical (full access with possible limitations, for team members); (iii) Advisory Board (limited access, for members of the advisory board who may need to access some specific documents).

The table below summarizes the accessibility of data described in Section 1 ("Data Summary").

Data	Dissemination Level	Deposition Location	SW tools for accession
Public Research Data	Public	Loughborough University's research data repository or Zenodo	To be defined in future versions of the DMP
Confidential Research Data	Confidential	Shared space on Loughborough University's servers	To be defined in future versions of the DMP
Project deliverables	Public/Confidential	Shared space on Loughborough University's servers; public deliverables will also be available for download on project website	Accessible with standard office SW
Promotional materials and templates	Public/Confidential	Shared space on Loughborough University's servers; promotional materials will also be available for download on project website	Accessible with standard office SW
Official documents	Confidential	Shared space on Loughborough University's servers	Accessible with standard office SW
Basic personal data	Confidential	Shared space on Loughborough University's servers	Accessible with standard office SW
Meeting data	Confidential	Shared space on Loughborough University's servers	Accessible with standard office SW

Table 4. Accessibility of data

3.3 Making data interoperable

The public research data produced during the project will mainly consist of datasets from experimental/simulation activities. Formats for such data is not fully established yet but will likely be standardised (e.g. .csv, .txt, etc.). However, due to the nature of the project, it is unlikely that such data will allow re-combination with different datasets.

More details on approaches for making data interoperable will be provided in future versions of the DMP.

3.4 Increase data re-use

All public research data will be made available simultaneously or shortly after publication by uploading on Loughborough University's research data repository. No licensing scheme has been established yet. The data will remain re-usable according to Loughborough University's research data repository policies (at least 15 years).

More details on approaches for increasing data re-use and licensing will be provided in future versions of the DMP.

4. Allocation of Resources

Estimation of costs for data management is problematic at this initial stage, considering the expected size of the data produced is not completely established yet. Data stored on Loughborough University's servers will be maintained without direct costs for the project.

Each partner will be responsible for uploading data (both public and confidential) on the shared space provided by Loughborough University according to project needs. Also, each partner will provide the data which will be made publicly available (i.e. Public Research Data, see Section 2 "Data Summary") to the project coordinator (ELTE) who is also the partner responsible for WP8 "Dissemination, Exploitation, Communication and Outreach", who will then deposit it on Loughborough University's research data repository.

5. Data Security

Partners involved in the research activities are responsible for providing data security and storing and creating regular backups of the data they are producing and/or processing according to the regulations of their organisations and usual practices (e.g. USB devices, backups in a central data centre, etc.). Data uploaded on Loughborough University's research data repository will be securely preserved according to the preservation policy of the repository. Security and preservation of data uploaded by partners on the shared space on Loughborough University's servers will be provided according to the regulations and usual practices of Loughborough University. Partners will exchange information via email, but confidential data should be exchanged exclusively through the shared space on Loughborough University's servers. Access to the shared space on Loughborough University's servers will be regulated using different authorisation levels (see Section 3.2, "Making data openly accessible"). At the end of the project, the shared space on Loughborough University's servers will be closed and all data will be permanently erased. Confidential data which might be useful in the future will be archived and preserved by partners according to the policies of QUEST Consortium Agreement.

6. Ethical Issues

Ethical issues are not expected to have any impact on the project.