D1.3: Initial Data Management Plan

Revision: v.0.7

Work package	WP1
Task	Task 1.4
Due date	03/06/2024
Submission date	17/06/2024
Deliverable lead	MART
Version	1.0
Authors	Eva Hajdok *(MART),* Amjad Yousef Majid (*MART)*
Reviewers	Davide Bacco *(QTI SRL),* E. Semenova *(DTU)*
Abstract	This document (to be updated at the end of the reporting period) will identify the best practices and specific standards for the generated data and assess their suitability for sharing and reuse in accordance with official EC guidelines
Keywords	QPIC-1550, Data management plan, Quantum Dots, Quantum Photonics, Quantum Key Distribution, Remote Quantum Computing, and Quantum Clock Synchronisation experiments

Grant Agreement No.: 101135785
EMERGING
Call: HORIZON-CL4-2023-DIGITAL-EMERGING-01-CNECT

Topic: HORIZON-CL4-2023-DIGITAL-

Type of action: HORIZON-RIA

Document Revision History

Version	Date	Description of change	List of contributors
V0.1	21/05/2024	1st edition of the data management plan	Eva Hajdok (*MART*)
V0.2	22/05/2024	Scientific revision and questionnaire editing	Amjad Yousef Majid *(MART)*
V0.3	06/06/2024	Incorporation of partners' input through the DMP questionnaire to the deliverable	Eva Hajdok (*MART*), Amjad Yousef Majid *(MART)*
V0.4	07/06/2024	Editing and formatting	Paola Ando *(MART)*
V0.5	12/06/2024	Revision by the peer-reviewers	Davide Bacco *(QTI SRL),* E. Semenova *(DTU)*
V0.6	13/06/2024	Incorporation of the comments by the peer-reviewers	Eva Hajdok (*MART*)
V0.7	17/06/2024	Final editing and formatting	Christos Tselebis *(MART)*

DISCLAIMER

Funded by the European Union

The information, documentation and figures available in this deliverable are written by the QPIC 1550 (Quantum photonic integrated circuits at 1550 nm) project's consortium under EC grant agreement 101135785 and do not necessarily reflect the views of the European Commission. The European Commission is not liable for any use that may be made of the information contained herein.

COPYRIGHT NOTICE

© 2023 - 2027 QPIC 1550

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

Classified C-UE/ EU-C	*EU CONFIDENTIAL under the Commission Decision No2015/ 444*	
Classified S-UE/ EU-S	*EU SECRET under the Commission Decision No2015/ 444*	

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

DEM: Demonstrator, pilot, prototype, plan designs

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

DATA: Data sets, microdata, etc.

DMP: Data management plan

ETHICS: Deliverables related to ethics issues.

SECURITY: Deliverables related to security issues

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

© 2023-2027 QPIC 1550

EXECUTIVE SUMMARY

The goal of this project is to create a universal QPIC platform at 1550nm by integrating InP, Quantum Dots and InGaAs detectors on SiN PICs together for the first time. This new platform has the potential to revolutionise quantum photonics by enabling the creation of on-demand single photons and entangled photon pairs, with efficient single photon detection. Critical to this platform its compatibility with existing optical fibre networks which makes this a cost-effective solution for quantum communication, computing, and quantum metrology applications. We demonstrate the flexibility of our QPIC platform in three different use cases: Quantum Key Distribution, Remote Quantum Computing and Quantum Clock Synchronisation experiments. The Data Management Plan (DMP) will ensure sound handling of data protection rules and describe the data management for the data to be collected, processed and/or generated and contain information on: the handling of research data during and after the end of the project; what data will be collected, processed and/or generated ; which methodology and standards will be applied; whether data will be shared/made open access ; how data will be curated and preserved (including after the end of the project), all with compliance with GDPR regulations.

Funded by the European Union

© 2023-2027 QPIC 1550

TABLE OF CONTENTS

© 2023-2027 QPIC 1550

ABBREVIATION

PDF	Portable Document Format
CSV	Comma Separated Values
DMP	Data Management Plan
JSON	Javascript Object Notation
GDPR	General Data Protection Regulation
JPG/JPEG	Joint Photographic Experts Group
MP4	Moving Picture Expert Group-4
WAV	Waveform Audio File
MP3	Moving Picture Experts Group Layer-3 Audio
PDK	Process Design Kit

Funded by the European Union

© 2023-2027 QPIC 1550

1 INTRODUCTION

This document outlines QPIC 1550 data management plan (DMP). It has been prepared in agreement with all project partners. This is the first version of the QPIC 1550 DMP. This version can change to reflect changes in the law such as changes in Regulation (EU) 2016/679 of the European Parliament and the Council of 27 April 2016 on the protection of natural persons about the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (General Data Protection Regulation — GDPR). Further, this document will be updated after each reporting period of the QPIC 1550 project in a form of 'D1.4 Interim review of the DMP' and 'D1.5 Final version of the Data Management Plan'. It will also be updated to reflect significant changes in the project such as changes in consortium policies, consortium composition, and new data.

Funded by the European Union

© 2023-2027 QPIC 1550

2 DATA SUMMARY

Beyond research publications, QPIC 1550 will collect and generate a rich variety and considerable amount of data and research outputs, which will be managed in line with the FAIR (Findable, Accessible, Interoperable, Reusable) principles. This data will serve three main purposes: research, communication and dissemination, and internal organisation. Scientific reports and data will be open access.

2.1 WHY GENERATING/COLLECTING DATA

Scientific reports will be generated to inform scientific communities about the breakthroughs and scientific advancements that are made during the development of QPIC 1550. Word documents will be used to track the progress of the work packages. Excel sheets will be used for internal organisation purposes. Flyers, social media posts, and e-newsletters will be generated to increase awareness about the objectives and progress of QPIC 1550.

It's important to note that the data is directly provided by the data subjects through reliable platforms and tools such as Google Forms and Excel Sheets. Each partner has its own internal data storage policy and protection measures in place, including network shares and dedicated Microsoft SharePoint or Google Drive servers. The folder structure on these servers is carefully defined and linked to frameworks or contracts, with access rights managed by the project manager for each collaborator based on their role.

2.2 DATA TYPES, FORMATS, AND ORIGIN

QPIC 1550 will generate different types of data such as deliverables, prototypes, and stimuli prepared for pilot actions, collected metrics including structured datasets, and developed software.

QPIC 1550 scientific publications will be in a PDF format. Word documents and Excel sheets will be used to track the progress and for internal organisation purposes. PNG/JPG images, MP4/MKV videos, and MP3/WAV audio files can be generated for social media and online posts. Data related to research activities can be generated in JSON, YAML, CSV, and binary format and stored in trusted OpenAire-compliant repositories.

Table 1: Data origin, format, type, and purpose

Data set	Personal	Origin	Format	Purpose
Stakeholder	Yes	self-registration to events and the QPIC 1550 website providing consent	.xls .json .docx	facilitate communication and dissemination activities

© 2023-2027 QPIC 1550

Stakeholder community contact data	Yes	self-registration to events and the QPIC 1550 website providing consent	.xls .json .docx	community landscaping, processing
Organisation profiles	Yes	self-registration of interest on the QPIC 1550 website, contribution to surveys	.csv	community landscaping, processing

2.3 DATA UTILITY

To effectively carry out project activities, it's important to make use of the collected data in a meaningful way, resulting in improved customer experience, loyalty, and engagement. The project partners will utilise the data for the following three main purposes:

1. To track project progress through participation in different QPIC 1550 events (both online and on-site), workshops, and training.
2. For further research
3. For dissemination to newsletter subscribers, future events, community members, and website visitors.

The data can be stored in partners' cloud infrastructures such as private Google Drive or OneDrive, according to the internal rules of the partner. Only authorised users will have access to the data through authentication.

2.4 RE-USING EXISITING DATA

QPIC 1550 will be re-using package designs and processes developed in another projects, adapted to the requirements of the QPIC 1550 project to deliver the desired performance of the use-case demonstrators. TUE has developed a PDK (process design kit) for the technology that will be used (and adapted) in QPIC1550. The PDK consists of a spec sheet that includes anticipated performance and gds (mask files) to make mask design layouts. Furthermore, the process is described in a PDK. Furthermore, TUE will use existing process recipes for etching, deposition, lithography, etc. These are cleanroom-specific but can typically be used in an open way by other cleanroom users, internal and external.

© 2023-2027 QPIC 1550

2.5 DATA SHARING WITH OTHER PARTNERS

In general, the partners will be sharing images, presentations and measurements data when specific tasks of the project require and during the internal meetings with selected partners upon mutual agreement. More specifically Tyndall will be working with other partners to make sure the prototypes align with their specifications and system requirements. This will include primary 3D CAD designs, packaging design rules as well as measured performance data, such as coupling efficiency.

Furthermore, UNIWA will most probably share data with DTU, QTI, Polimi, Tyndall related to the final applications on distributed computing and metrology, as long as designs of chips (1st and 2nd generation) In addition, the PDK will be shared with other partners by TUE, so they can design in our TUE process. TUE will then fabricate the wafers. The PDK will contain all relevant information for the partners to make a design. This can still mean that proprietary building blocks (components) are given as a so-called black box, i.e., only performance data are given and footprint, but no process and layout information. This is a typical way of sharing a PDK in the semicon world. POLIMI will also share the required data with other partners when the project reports will have to be prepared. Additionally, during the execution of the project POLIMI will share the minimum amount of data with other partners for the proper collaboration in the project development.

© 2023-2027 QPIC 1550

3 FAIR DATA

QPIC 1550 follows the principles of FAIR data management (Findability, Accessibility, Interoperability, and Reuse). The aim of these principles is to make digital assets more easily discoverable, accessible, interoperable, and reusable for researchers and practitioners. By following the FAIR principles, data becomes more valuable and usable, enabling better collaboration and knowledge sharing.

The measures for FAIR data management include:

- **Findability** includes the use of standards and metadata format, persistent and unique identifiers, naming conventions, search keywords, and version numbers, which help to make data discoverable and accessible. Findability ensures that data can be easily located and retrieved, allowing others to easily access, reuse, and build upon it.

- **Accessibility** refers to the ease with which data can be located and retrieved by both humans and machines. It involves making data available and usable to anyone who needs it, without any barriers such as technical, legal, or institutional restrictions. This includes making data available in a machine-readable format and providing adequate documentation to understand and use the data. Accessibility also requires data to be properly stored and maintained, so that it remains usable over time. By ensuring accessibility, FAIR data management helps to promote data reuse and scientific collaboration.

- **Interoperability** refers to the ability of different systems, platforms, and tools to work together and exchange data effectively. This means that data should be stored, formatted, and described in such a way that it can be understood and used by different systems and applications, regardless of the technology used to create or store the data. Interoperability is an important aspect of FAIR data management because it enables seamless data exchange between different systems and facilitates data reuse, which is crucial for scientific collaboration and discovery. Interoperability helps to ensure that data is accessible, usable, and valuable over time, and it supports the efficient integration of new data and tools into existing systems.

- **Reusability:** Reusability in fair data management refers to the ability to use data in multiple ways, for multiple purposes, with proper attribution and consent, while ensuring that the data remains secure and respects the privacy of individuals. This principle is an important aspect of fair data management as it helps to maximise the value of data, promote scientific progress, and support innovation while still respecting ethical considerations such as privacy and confidentiality.

- **Resource allocation:** costs of making data FAIR, how to cover these costs, the role of the data manager, partner responsibilities, the potential value of long-term preservation, and preservation costs.

© 2023-2027 QPIC 1550

- **Archiving and preservation:** data at the end of the project, data selection for archiving and preservation, estimated final volume, recommended preservation duration, and long-term preservation storage.

- **Data security:** data security provisions and security of long-term preservation.

- **Ethical considerations**: Impact of ethical or legal issues. To ensure a user-friendly approach to the preparation and maintenance of the data. The DMP will be stored in the Project Platform (SharePoint) where it will be accessible to all partners.

QPIC 1550 will integrate data from all applications participating in the community, website (newsletter subscribers), events (event registrants), interviews, focus groups, workshops, and training. The data will be collected through an online application form, which will standardise user information and facilitate data integration. The data management plan (DMP) will be stored on the Project Platform (SharePoint) and accessible to all partners.

3.1 MAKING DATA AND METADATA FINDABLE

QPIC 1550 data and research outputs will be stored and appropriately catalogued in trusted OpenAire-compliant repositories. Zenodo (https://zenodo.org) will be used by default and a specific section will be created and linked to the project. These outputs will be assigned persistent and unique identifiers (e.g., digital object identifiers) and metadata (e.g. ids, keywords, data type, authorship, licenses, version…) to maximise their findability and identification through search engines. Links to the repositories in the project web page will foster the findability and accessibility of the data.

Rich metadata refers to the complete and detailed information that is associated with a dataset or digital object. This information goes beyond basic metadata (such as file name, date created, and file size) to include information that provides context and meaning to the data. This can include information such as the data's origin, purpose, quality, provenance, and relationships to other data. Rich metadata also includes descriptions of the data's structure, format, and any relevant processing or analysis that has been performed on it. Such information enables effective discovery, reuse, and management of data, as it provides the information necessary for others to understand and use the data. **QPIC 1550 data will be accompanied by rich metadata.** For example,

1. All project-related documents must include the author(s) & contributor(s), WP, dissemination level, document nature, synopsis, and keywords on the front page.

2. For information submitted by registered users, the criteria listed in the application form will also be used to identify and make the data discoverable. The registration form uses multiple-choice questions to create a database and identify users based on their characteristics

3. The readme files and associated webpages of the containerized applications, databases, and machine learning models will have data description, data type, keywords, authorship, licenses, version, and publication date.

© 2023-2027 QPIC 1550

4. For scientific publications the metadata must at least include authors names, title, date of publication, publication venue. Also, Horizon Europe or Euratom funding; grant project name, acronym and number; licensing terms; persistent identifiers for the publication, the authors involved in the action. Where applicable, the metadata must include persistent identifiers for any research output or any other tools and instruments needed to validate the conclusions of the publication.

The information submitted by users via the registration forms will use keywords related to QPIC 1550 technologies such as Quantum Dots, Quantum Light Sources, Quantum Photonics, Quantum Key Distribution, Remote Quantum Computing, and Quantum Clock Synchronisation experiments. These keywords will also be used to easily identify users based on their specific interests during communication activities.

An excel spreadsheet will be created to catalogue information about users and determine the size of interested parties within a specific area. Excel will serve as an efficient tool for filtering users by characteristics and improving community communication. The export functionality of each partner's platform will enable this list to be exported.

By this QPIC 1550 adheres to the FAIR data principles (Wilkinson et al., 2016) which states that (meta)data should (i) Have a globally unique and permanent identifier; (ii) include sufficient metadata to fully comprehend the data; and (iii) Be indexed in a searchable repository. These principles ensure the data is retrievable and includes authentication and authorization information.

3.2 MAKING DATA ACCESSIBLE

QPIC 1550 commits to bringing research results closer to the public and adhering to the Open Access guidelines set by the Horizon Europe work programme. In line with these guidelines, all the scientific publications supported by the project will be available as Open Access through an OpenAire-compliant repository such as Zenodo and the Open Research Europe publishing platform, allowing for easy linking with the EU-funded project. This will increase the accessibility to the obtained results by a wider community, which can be further enhanced by including the repository in the registries of scientific journals.

Research data will be published upon publishing the associated research results/report. Embargoes may be placed on theses and/or data to protect intellectual property, but once legally protected, embargoes will be lifted. The data will be accessible via free and standardised tools such as web browsers. In addition to that, research-related data will be made accessible programmatically to facilitate future development.

There is no predetermined limit on how long the data will be available and findable. In other words, we will maintain data accessibility as long as possible. The code and API documentation will be open access. However, a beneficiary that intends to disseminate its results must give at least 15 days advance notice to the other beneficiaries (unless agreed otherwise), together with sufficient information on the results it will disseminate. Any other beneficiary may object within (unless agreed otherwise) 15 days of receiving notification, if it can show that its legitimate interests in relation to the results or background would be

© 2023-2027 QPIC 1550

significantly harmed. In such cases, the results may not be disseminated unless appropriate steps are taken to safeguard those interests.

Regarding the peer-reviewed scientific publications, the beneficiaries shall ensure open access to their results. They will ensure that:

at the latest at the time of publication, a machine-readable electronic copy of the published version or the final peer-reviewed manuscript accepted for publication, is deposited in a trusted repository for scientific publications immediate open access is provided to the deposited publication via the repository, under the latest available version of the Creative Commons Attribution International Public Licence (CC BY) or a licence with equivalent rights; for monographs and other long-text formats, the licence may exclude commercial uses and derivative works (e.g. CC BY-NC, CC BY-ND) and information is given via the repository about any research output or any other tools and instruments needed to validate the conclusions of the scientific publication. Beneficiaries (or authors) must retain. Finally, all metadata will be made openly available and licensed under a public domain dedication CC0 or equivalent as stated in the Grant Agreement.

3.3 MAKING DATA INTEROPERABLE

Data interoperability refers to the ability of different systems to exchange and use data in a seamless and mutually understood manner. QPIC 1550 will strive to make the data interoperable by adopting/following the following steps.

- Adopt common data formats: Choose data formats that are widely accepted and used, such as CSV, JSON, or XML, to ensure that the data can be easily understood and processed by a variety of systems

- Use common data standards: Utilise data standards, such as ISO or W3C standards, to ensure that data is structured and labelled consistently, making it easier for different systems to understand and process the data.

- Document data structure and content: Provide clear and comprehensive documentation of the data structure and content, including the meaning and use of each field, to help ensure that data is understood and used consistently.

- Use APIs: Provide APIs (Application Programming Interfaces) to allow different systems to easily access and exchange data, without the need for manual data conversion or manipulation.

- Establish data governance practices: Implement data governance practices to ensure that data is managed consistently, with clear roles and responsibilities, and that data quality is maintained over time.

© 2023-2027 QPIC 1550

3.4 DATA REUSABILITY AND QUALITY

Online data and code will be accompanied by documentation and readme files to facilitate data reuse. To ensure the widest re-useability, the data, and code will be freely available in the public domain under standard reuse licences in line with the obligations set out in the Grant Agreement. Third-party users will be encouraged to use the system and data produced under the QPIC 1550 project. Finally, the provenance of the data will be documented using appropriate standards. A more detailed data reusability plan will be provided in the next version of this document.

The project coordinators will be responsible for assuring the quality of the data by making sure the dataset follows the FAIR principles included in this plan, and it is up to date.

3.5 COST OF MAKING THE DATA FAIR

No extra costs are expected to make the data FAIR. The project coordinators will be responsible for maintaining and updating the data management plan.

© 2023-2027 QPIC 1550

4 DATA SECURITY AND PRIVACY

The utmost care should be taken when processing data and strict compliance with established procedures must be maintained. Many of these procedures are mandated by European law and regulations, including the General Data Protection Regulation (GDPR). Data generated/collected under the umbrella of QPIC 1550 will undergo a strict ethical review process that strictly adheres to the laws and regulations. The data will be stored in safe and secure repositories such as Microsoft SharePoint, Google Drive, and GitHub to ensure long-term preservation and curation.

The personal data collected for the QPIC 1550 project will come mainly from the sign-up form submission. In addition to the sign-up form, for big events a paper form can be used to ensure the registration of everyone. Before collecting data from participants, they will be provided with detailed information about the data processing and how this process adheres to Articles 13 and 14 of the GDPR. This will allow participants to make informed decisions about providing their data and participating in the project. The GDPR requirement ensures that participants have control over their personal data.

Funded by the European Union

© 2023-2027 QPIC 1550

5 ETHICS

The QPIC 1550 Consortium is committed to perform the project in compliance with the General Data Protection Regulation 2016/679 ("GDPR") and any implementing local legislation (collectively referred to as the "EU Data Protection Legislation") before making any resource and data openly accessible. To this end, the partners will ensure that no personal data (as such term is defined in the GDPR) will be shared between the Partners unless (i) it has been fully anonymised prior to the data sharing, or (2) the specific partners who have elected to exchange or otherwise process personal data, have entered into separate data processing agreement and have determined what operational measures should be taken prior to such personal data exchange or processing, all in accordance with the EU Data Protection Legislation. Further, no aspect of the project should force a party to violate any mandatory laws or regulations, including those related to ethics in research.

QPIC 1550 participants will be fully informed about the intended use of the information collected from them and must agree to the data collection for the specific purpose with their active approval in the form of written consent. An Informed Consent Form will be handed out to any individual participating in QPIC 1550 activities.

Regarding the project website, and the communication tools that will be used in the project, any personal data collected (newsletters registrations) will be used only for outreach purposes and only in the context of QPIC 1550. Any other data collected during events, workshops, and webinars that are planned within the project might regard personal data, sociodemographic, and professional data and they will be made available only with the consent of the participants. Any analysis and use of these data will be after anonymisation and only for the purposes of the project.

When it comes to the 3rd party projects a process to ensure that the initiatives and projects that are funded comply with the ethical standards and guidelines of Horizon EU will be set up (Task 10.3 of the work plan). An ethical assessment will be done on all selected third-party projects by external experts according to the ethical guidelines of Horizon EU. The resulting report may include specific requirement (contractual obligation) in the ethics section of the Sub-Grant Agreement.

Horizon EU rules on ethical issues will be followed and, in case of conflict, the Project Coordinator and the relevant project management board will be informed and will solve the issue following national legislation and EU Directives. All this will guarantee that the ethical standards and guidelines of Horizon EU will be rigorously applied, regardless of the country in which the project activities are carried out.

The beneficiaries will take all measures to promote equal opportunities between men and women in the implementation of the action and, where applicable, in line with the gender equality plan. They will aim, to the extent possible, for a gender balance at all levels of personnel assigned to the action, including at supervisory and managerial level.

Finally, QPIC 1550 will not make use of other national, funder, sectorial, or departmental procedures for data management.

Funded by the European Union

© 2023-2027 QPIC 1550

6 CONCLUSIONS

As described at the beginning of the document The QPIC 1550 Data Management Plan provides detailed information on the procedures that will be implemented for data collection, storage, protection, retention, and destruction and confirmation that they comply with national and EU legislation, all along the project implementation.

The DMP (current version) is taking into account the requirements of the General Data Protection Regulation (GDPR), which is applicable as of May 25th 2018, and has introduced multiple changes with respect to the former legislation (i.e. Directive 95/46/EC) regarding 'protection of the fundamental rights and freedoms of natural persons and in particular, their right to the protection of personal data, and Directive 2002/58/EC on privacy and electronic communications (M6). Further, this document will be updated after each reporting period of the QPIC 1550 project in a form of 'D1.4 Interim review of the DMP' and 'D1.5 Final version of the Data Management Plan'. It will also be updated to reflect significant changes in the project such as changes in consortium policies, consortium composition, and new data.

Funded by the European Union

© 2023-2027 QPIC 1550