Where Quantum and Fluids entangle

Quantum Computational Fluid Dynamics

Data Management Plan

Project number# 101080085

Call:	HORIZON-CL4-2021-DIGITAL-EMERGING-02
Topic:	HORIZON-CL4-2021-DIGITAL-EMERGING-02-10
Type of action:	HORIZON Research and Innovation Actions
Granting authority:	European Commission-EU
Project starting date: fixed date:	1 November 2022
Project end date:	31 October 2026
Project duration:	48 months
Project Coordinator:	University of Hamburg (UHH)
Deliverable number:	D10.1
WP contributing to the deliverable:	WP#10, Project and Data Management, Dissemination and Communication
Deliverable Type:	Data Management Plan
Revision:	1
Dissemination level:	Public
Due Submission date:	18.04.2024
Prepared By:	Oscar Negrete, Bodo Krause-Kyora, Belda Atilla
Internal Reviewers:	
Final Approval:	Prof. Dr. Dieter Jaksch

This project receives funding from the European Union's Horizon 2020 HORIZON Research and Innovation Actions Programme under Grant Agreement #101080085

Revision History

Version	Date	Who	Changes
0	14.12.2023	Oscar Negrete	Created Document
1	18.04.2024	Belda Atilla	The document was rejected due to the amendment process. Revision changed and re-submitted.

Executive Summary

The Quantum Computational Fluid Dynamics (QCFD) project is dedicated to establishing an open-access quantum software framework to address widespread challenges within computational fluid dynamics present in today's industry. This Data Management Plan (DMP) delineates the protocols for managing the vast array of data generated by the project. Central to our dissemination strategy is the establishment of the web portal www.qcfd-h2020.eu that will act as a centralized repository, facilitating access to project-related datasets, benchmark results, and software tools.

In adherence to the FAIR data principles — Findability, Accessibility, Interoperability, and Reusability — our approach encompasses the provision of a detailed metadata structure for each dataset. This will include unique identifiers and proper documentation, ensuring that the data is not only easily accessible but also readily transferable and reusable. This system will be implemented on a dedicated storage server hosted at the University of Hamburg, which has the necessary security, backup, and data restoration protocols to ensuring the safekeeping of our datasets and facilitating consistent access to our online platform.

Throughout the lifespan of the QCFD project, this Data Management Plan (DMP) will be subject to regular updates at intervals of three to six months. Each update will aim to incorporate specific details and refinements for each protocol involved in the management and handling of data.

Contents

This project receives funding from the European Union's Horizon 2020 HORIZON Research and Innovation Actions Programme under Grant Agreement #101080085

List of abbreviations

Acronym / Short Name	Meaning
CA	Consortium Agreement
DMP	Data Management Plan
EC	European Commission
ENG	ENGYS SRL
EU	European Union
FZL	Jülich Research Centre
KPI	Key Performance Indicator
PlanQc	PlanQc GMBH
PO	Project Officer
QCFD	Quantum Computational Fluid Dynamics
TUHH	Technical University of Hamburg
TUC	Technical University of Crete
TUM	Technical University of Munich
UHH	University of Hamburg
WP	Work Package
WPL	Work Package Leader

1. Introduction

1.1. Definitions

Access rights:	Rights to use results or background.
Dissemination:	The public disclosure of the results by appropriate means, other than resulting from protecting or exploiting the results, including by scientific publications in any medium.
Exploit(ation):	The use of results in further research and innovation activities other than those covered by the action concerned, including among other things, commercial exploitation such as developing, creating, manufacturing and marketing a product or process, creating and providing a service, or in standardisation activities.
Fair and reasonable conditions:	Appropriate conditions, including possible financial terms or royalty-free conditions, taking into account the specific circumstances of the request for access, for example the actual or potential value of the results or background to which access is requested and/or the scope, duration or other characteristics of the exploitation envisaged.
FAIR principles:	'findability', 'accessibility', 'interoperability' and 'reusability'.
Open access:	Online access to research outputs provided free of charge to the end-user.
Open science:	An approach to the scientific process based on open cooperative work, tools and diffusing knowledge.
Research data management:	The process within the research lifecycle that includes the organisation, storage, preservation, security, quality assurance, allocation of persistent identifiers (PIDs) and rules and procedures for sharing of data including licensing.
Research outputs:	Results to which access can be given in the form of scientific publications, data or other engineered results and processes such as software, algorithms, protocols, models, workflows and electronic notebooks.

1.2. Purpose of the Induce Data Management Plan (DMP)

The EU Framework Programme for Research and Innovation Horizon 2020 aims to improve and maximise access to and re-use of research data.

Data Management Plans (DMPs) are a key element of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project. As part of making research data findable, accessible, interoperable and re-usable (FAIR), a DMP should include 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 and

- how data will be curated and preserved (including after the end of the project).

This document has been produced following these guidelines and aims to provide a consolidated plan for QCFD partners in the data management plan policy that the project will follow. Regular updates to the DMP will be implemented to keep pace with the project's progression and emerging data management trends.

1.3. Background of the Induce DMP

The QCFD Data Management Plan (DMP) is designed in alignment with the stipulation outlined in Article 16 of the Grant Agreement, which pertains to Intellectual Property Rights (IPS), encompassing Background and Results, Access Rights, and Rights of Use. Additionally, it adheres to the specifications defined in Article 17 (Communication, Dissemination, and Visibility).

This document presents the initial iteration of the Data Management Plan (DMP) specifically devised for the QCFD Project. It delineates a comprehensive framework encompassing the policies and strategies for managing the diverse datasets produced during the project's lifecycle. Specifically, it articulates the methodologies for the systematic collection, generation, processing, and dissemination of data. Additionally, it elaborates on the protocols for the secure and efficient disclosure and storage of the project's data assets.

2. Data Summary

The QCFD project's data generation is oriented to the development of numerical scientific libraries. This endeavor is set to culminate in a computational framework that seamlessly integrates both classical and quantum numerical methods. Furthermore, the data will provide essential guidelines to facilitate the strategic management and coordination of the project, research directions, administrative operations, and the strategic dissemination of scientific findings.

The core of the dataset will emanate from computational simulations derived from both internal and external collaborations within the QCFD partners. Data classification will be carefully implemented, distinguishing between confidential and non-confidential (public) categories in adherence to the project's regulatory basis. This division is crucial for safeguarding the intellectual property inherent in the scientific developments of the project while supporting scientific collaboration.

The formats and type of data generated in QCFD will contemplate the representation appropriate to the tasks related to management, scientific development, and dissemination. Within the context of numerical data management, the results will be stored mainly in HDF, CSV, and JSON formats, which provides a wide comparability and transferability between industry and academia. Reporting, analysis, and logging will be handled largely within the Microsoft ecosystem, as will much of the dissemination of scientific results. Additionally, the dissemination process will make extensive use of multimedia formats (videos, images, and audio).

The impact of the data generated by QCFD is expected to significantly contribute both during and post-project. Its value is anticipated to extend beyond the project's tenure, offering tangible benefits to industrial stakeholders, the scientific community, and society, with varying degrees according to specific goals and interests.

3. FAIR data

3.1. Making data findable, including provisions for metadata

The data within the scope of the QCFD project will be securely stored in the PHYSnet data center, located at the University of Hamburg. Access and retrieval of this data are governed by a well-defined access management protocol, adhering to the center's security and backup policies, as delineated in subsequent sections. The principal access to software development data will be facilitated through PHYSnet's proprietary version control system, designed to maintain the integrity and versioning of the software development datasets.

To address the technical nature of our data, comprehensive documentation will be provided to guarantee clarity in utilization and interpretation. This is crucial for transparent replication of our methodologies, promoting accessibility for scientific and industrial sectors. The metadata accompanying simulation datasets will be meticulously detailed, including parameters, software versions, library dependencies, and simulation timeframes.

To enhance the findability of our data/datasets, a keyword-based search system supported by Digital Object Identifiers (DOIs) will be deployed. This system is designed to provide seamless access via our dedicated platform, as well as through conventional web search engines, thereby increasing the visibility and navigability of our databases.

The majority of data produced by the QCFD project will be openly accessible to both the academic community and the general public, adhering to open-source principles. In specific instances where data is intended solely for academic purposes, options for its non-academic use and potential commercialization will be open for discussion, in line with the project's internal guidelines. This approach ensures compliance with intellectual property rights, balancing open access with responsible data utilization.

3.2. Making data openly accessible

To enhance data accessibility in the QCFD project, efforts will be made to ensure that proprietary QCFD data formats and structures are as compatible as possible with widely used open-source tools. General datasets, along with comprehensive documentation, will be made directly accessible via the project's web platform. For more specialized technical data, access will be facilitated through PHYSnet's version control system, available at the PHYSnet Git Repository (https://git.physnet.uni-hamburg.de/).

On the web platform, detailed instructions will be provided to articulate the overall structure of data management within the project. This will include guidance on accessing datasets, libraries, and software, as well as information on the open-source tools necessary for data retrieval and analysis. To ensure transparency, the access conditions for each dataset will be displayed, accompanied by the appropriate licenses that delineate the separation between public and private data.

For access to sensitive or protected data, a clearly defined process will be established. Requests will require the submission of an intent questionnaire, which will be reviewed by a dedicated committee comprising representatives from QCFD and its industrial partners. This committee will ensure that access

to protected data is granted in a manner that respects the integrity of the data and the proprietary rights of project contributors.

Additionally, we will explore the implementation of an automated access system for non-confidential datasets, which would streamline the process for users and reduce administrative overhead. By employing an API for data retrieval, users could access public datasets programmatically, thereby improving the utility and integration of QCFD data with external applications and services.

The QCFD project will regularly review and update its accessibility protocols to align with emerging standards and community best practices. Our goal is to maintain an open, transparent, and user-friendly data environment while upholding the standards of data protection and intellectual property rights.

3.3. Making data interoperable

To facilitate rapid acquisition of the tools and data generated in QCFD, the terminology used in the metadata for labeling data will align with the prevailing lexicon within both the industrial and academic communities. This standardization will extend to the process of enhancing the compatibility of proprietary formats with open-source tools, ensuring a seamless integration of data across various platforms.

For those instances where the QCFD project introduces specialized terminology or unique data structures that are not prevalent in current literature, we will provide detailed documentation. This documentation will aim to elucidate these novel concepts and structures, ensuring they are comprehensible and usable beyond the immediate scope of the project.

Recognizing the interdisciplinary essence of QCFD and the diversity of its industrial partners, there is an ongoing initiative to develop comprehensive instructions that foster cross-departmental communication. These instructions will be designed to bridge knowledge between different sectors involved in the project, presenting each aspect from multiple perspectives to enhance understanding and collaboration.

The QCFD project will be actively engaging in regular dialogues with industry experts and academic peers. These exchanges will enable us to stay abreast of emerging trends and best practices in the field. Insights gained from these interactions will be instrumental in refining our interoperability strategies, thereby maintaining our data's relevance and compatibility with evolving standards and technologies prevalent in both industrial and academic sectors.

3.4. Increase data re-use (through clarifying licenses)

In alignment with our commitment to promoting scientific advancement and collaboration, the QCFD project will adopt an open-source approach for all datasets and software tools developed through a Creative Commons Attribution (CC BY-NC) license.

For specific scenarios that involve potential commercial utilization, such as the resale or distribution of our data to third parties, we will implement a commercial agreement according to the requirements of the entities involved. This agreement is intended to manage and regulate explicit commercial uses of our data, safeguarding the interests and intellectual property rights of the QCFD project and its industrial stakeholders. Commercial entities interested in such applications will need to engage in negotiations (commercial use on request) with

our project team. During these discussions, we will establish specific terms of use tailored to each case. The negotiation will cover various aspects, including licensing fees, the scope and duration of use, and any required safeguards for sensitive data or methodologies.

4. 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.).
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.

5. Allocation of resources

In the QCFD project, our approach to resource allocation for data storage is both strategic and cost-effective. We have chosen to utilize the servers hosted within the PHYSnet system at the University of Hamburg for our project's data storage needs. This decision utilizes the university's robust and secure infrastructure, a critical factor in ensuring the integrity and longevity of our research data. The University of Hamburg offers this service without imposing any maintenance charges. The costs associated with the upkeep of these servers are entirely covered by the university's internal policies.

Upon completion of the QCFD project, all generated and collected data will remain accessible for a minimum period of 5 with their respective licenses and access agreements. To ensure sustained availability, a dedicated long-term supervision team will be established. This team, in collaboration with the IT-PHYSnet staff, will develop and implement a comprehensive protocol designed to guarantee specialized access to the data in the subsequent years.

6. Data security

To ensure the integrity and availability of project data, a comprehensive backup strategy is implemented as follows:

The project's data is subject to daily mirroring processes. These operations are conducted in a dedicated backup server room, strategically located in a zone with reduced fire risk. Data transfer is facilitated through a robust 100 gigabit Ethernet connection to a directly connected storage system. This precautionary measure is designed to ensure that, in the event of a catastrophic failure, the impact is confined to the potential loss of only the most recent day's data, which can be rapidly restored to minimize downtime and data unavailability.

In conjunction with the daily mirroring, the OpenAFS system is configured to create a snapshot of each user's or project's volume every 24 hours. These snapshots are then mounted within the directory of the corresponding volume, granting users direct access to the previous day's data. This level of access autonomy allows users to perform data restoration operations independently, without necessitating administrative intervention.

For long-term data retention, a daily Tivoli Storage Manager (TSM) backup is executed, transferring data from the primary storage snapshots to tape robots. This process adheres to the 10-year retention guideline set forth by the Deutsche Forschungsgemeinschaft (DFG), providing a historical archive of project data that fulfills both regulatory compliance and research continuity requirements.

The integration of these backup services offers a multi-layered approach to data restoration, addressing various scenarios ranging from user error rectification to full-scale disaster recovery. The implementation of this strategy is critical for safeguarding the project's data against loss or corruption, ensuring operational resilience and data preservation in alignment with best practices and project governance policies.

7. Ethical aspects

The QCFD project has diligently assessed all ethical dimensions of our research and confirms that there are no ethical concerns. Our work, focusing on computational simulations and theoretical model development, does not involve human subjects or sensitive data. We adhere strictly to the ethical guidelines provided by the University of Hamburg, the FAIR data principles, and the ethical framework of Horizon 2020, ensuring compliance with all relevant regulations. Additionally, our data management practices fully respect intellectual property rights and ethical standards, ensuring integrity and transparency in our research.