

# QTCG Data Management Plan

## 1. Data Summary

Our Group's projects involve mostly, but are not limited to, Theoretical Physics and Quantum Information Science. The data lifecycle involves the following components:

- **Scientific Publications & Preprints:** Theoretical derivations and paper drafts (.tex, .pdf).
- **Source Code:** Python algorithms (.py), Jupyter Notebooks (.ipynb) for simulations, and quantum control scripts.
- **Research Data:**
  - **Simulation Data:** Numerical results generated via numerical simulations (typically .txt, .h5 or .csv).
  - **Quantum Data:** Output counts and probability distributions from quantum hardware/simulators (standard text formats).
  - **Analitical Data:** Calculations derived mostly from exact equations.

**Value and Sharing:** The data allows for the validation of the theoretical models and quantum algorithms proposed. All non-sensitive data, code, and resulting publications will be shared to facilitate reproducibility.

## 2. Responsibilities

- **Principal Investigator (PI):** The PI of each project is responsible for the overall DMP strategy and ensuring the team uses the designated platforms (GitHub, Zenodo, arXiv) correctly.
- **Research Team:**
  - Responsible for maintaining the group's **GitHub** repositories (committing code, documenting changes).
  - Responsible for uploading final datasets to **Zenodo** and preprints to **arXiv**.
- **Intellectual Property:** The code and data are generated under the rules of the University of Granada, but will be released under open licenses to maximize dissemination as detailed in Section 4.

### 3. FAIR Data Principles.

The data generated in our projects should fulfill the FAIR principles. Meaning that it should be:

#### A. Findable

- **Identifiers (DOIs):** We will use **Zenodo** to assign a Digital Object Identifier (DOI) to the final datasets and specific releases of the software code.
- **Preprints:** Drafts deposited in **arXiv** will be assigned a permanent arXiv ID (and subsequently a DOI).
- **Metadata:** We will utilize the standard metadata fields provided by Zenodo (DataCite schema. and Github to ensure rich description (authors, abstract, keywords, grant information).
- **Version Control:** All code development will be tracked using **Git** and hosted on the group's **GitHub** account, ensuring every version is traceable.

#### B. Accessible

- **Access Protocols:**
  - **Drafts:** Immediately accessible via **arXiv**.
  - **Code:** Publicly accessible via the Group's **GitHub** repository. We will use the GitHub-Zenodo integration to create persistent archives of the code at the time of publication.
  - **Raw Data:** Openly accessible via **Zenodo**.
- **Software Requirements:** A requirements.txt or environment.yml file will be included in the GitHub repository to define the Python environment (e.g., Qiskit, NumPy versions) needed to use the data.

#### C. Interoperable

- **Formats:** We prioritize non-proprietary formats:
  - **Data:** Plain text, HDF5 (hierarchical data) and JSON/CSV for maximum compatibility.
  - **Quantum Circuits:** We will use OpenQASM (an open standard for quantum instructions) where applicable.
- **Vocabularies:** Variable naming and metadata will follow standard Physics and Quantum Information conventions to ensure cross-compatibility with other research groups.

## D. Reusable

- **Licensing:**
  - **Code (GitHub):** Released under a permissive license (e.g., MIT or Apache 2.0) to encourage modification and reuse.
  - **Data (Zenodo):** Released under Creative Commons Attribution (CC-BY 4.0).
  - **Drafts (arXiv):** Released under the arXiv non-exclusive distribution license.
- **Embargo:** Generally, data and code will be made public simultaneously with the arXiv preprint submission. If a journal requires it, a temporary embargo may be set on Zenodo until acceptance.

## 4. Data Security

- **Active Storage:** Daily work is stored on institutional computers with backups. Code is continuously pushed to **GitHub**, which serves as a distributed off-site backup.
- **Long-Term Preservation:**
  - **Zenodo:** Guarantees long-term preservation of the data and code snapshots within the CERN Data Centre infrastructure.
  - **arXiv:** Ensures the perpetual availability of the scientific manuscript.
- **Ethics:** As this project involves theoretical physics and simulations, it does not involve personal data, human subjects, or animals. Therefore, no specific ethical security measures regarding privacy or sensitive data are required