

PQE Data Management Plan (DMP) Template

Project Information

Project title: [Insert title]

Principal investigator: [Name, department, institution]

Project dates: [Start date–end date]

Version/date: [DMP version and date]

1. Data Types and Metadata: Identify the types of data to be collected and the key metadata features associated with each

This project will generate and/or collect the following data:

Data Type	Description/source	Estimated Volume	Key Metadata
[e.g., Raw instrument]	[Describe data and collection method]	[Size]	[Sample ID, date/time, instrument, settings, operator]
[e.g., Processed/derived]	[Describe processing or analysis]	[Size]	[Processing steps, software/version, parameters, provenance]
[e.g., Documentation/code]	[Protocols, scripts, notebooks]	[Size]	[Author, date, version, license]

Metadata will be recorded using [file naming convention, electronic lab notebook, database, README files, or relevant community standard].

2. Quality Control and Processing: Describe the data quality control pipeline(s) and analysis/processing tools to be used

Data quality will be monitored through:

- [Calibration, duplicate measurements, controls, validation checks, or other procedures]
- [Automated checks for missing, invalid, or inconsistent values]
- [Manual review and correction procedures]
- [Version control and tracking of data-processing steps]

Data will be processed and analyzed using [software/tools, programming languages, and versions]. Processing workflows will be documented in [scripts, notebooks, workflow managers, or protocols] and maintained using [e.g., Git/version-control system]. Original raw data will be preserved and kept separate from processed or derived datasets.

3. Storage and Archival Needs: Estimate predicted dataset sizes and identify appropriate archival resources

The anticipated data volume is approximately:

- Raw data: [Estimated size]
- Processed data: [Estimated size]
- Code and documentation: [Estimated size]
- Total working and backup storage: [Estimated size, including redundancy]

During the project, data will be stored on [institutional server, approved cloud storage, secure workstation, or other resource]. Backups will be performed [frequency] using [backup system].

At project completion, data and documentation will be archived in [institutional repository, discipline-specific repository, or trusted general repository]. Data will be retained for at least [retention period] in accordance with institutional, funder, legal, and ethical requirements.

4. Publication Readiness: Anticipate required data formats and any open-access obligations associated with eventual manuscript submission

Data intended for publication or reuse will be organized and exported in non-proprietary or widely supported formats where possible, such as:

- Tabular data: .csv or .tsv
- Text and documentation: .txt, .md, or .pdf
- Images: [e.g., .tif, .png, or other appropriate format]
- Code: [e.g., .py, .R, or repository-specific format]
- Structured data: [e.g., .json, .xml, or discipline-specific format]

The final dataset will include [e.g., a README file, data dictionary, methodology, relevant code, processing documentation]. Data will be assigned a persistent identifier, [e.g., DOI, ARK, URN]

Any open-access requirements from [e.g., the funder, institution, or target journal] will be followed. Data will be deposited in an [appropriate repository] subject to [e.g., privacy, consent, intellectual-property, and confidentiality restrictions]. If full public release is not possible, the project will provide [e.g., restricted access, de-identified data, or appropriate metadata] and access instructions.