3. Data Management Plan (UKAEA-Aligned) *(Completed 10:30–10:40)*

This Data Management Plan (DMP) follows the basic structure recommended by **UKAEA** and **UKRI** for small-scale computational research projects. The plan outlines how simulation data, analysis scripts, and documentation are generated, stored, backed up, organised, and version-controlled throughout the project.

3.1 Data Types

The project will generate the following categories of data:

A. Code

- epic1d.py (Particle-in-Cell solver)
- scan_epic1d.py (automated parameter-scan tool)
- Additional analysis scripts (peak detection, noise estimation, frequency/damping tools)

B. Raw Simulation Outputs

Generated automatically by scripts:

- First-harmonic amplitude time series
- Peak locations and amplitudes
- Noise region data
- Repeated-run statistics
- Parameter scan results (CSV format)
- Execution time logs

C. Processed Data

- Computed ω, γ, noise level, RMS values
- Statistical averages ± uncertainties
- Convergence study results

D. Documentation and Reports

- Lab notebook (Google Docs)
- README files
- Slide presentations (ODP)
- Figures (PNG)

E. Metadata

Automatic metadata stored with each run:

- Date/time
- Simulation parameters (Np, Nc, L)
- Script version
- Random seed (optional)

3.2 Storage and Access

Local Storage

Data is stored in a structured directory:

```
lab/
  src/
    epic1d.py
    scan_epic1d.py
  results/
    scan_runs/
      csv/
      logs/
      plots/
    single_runs/
      harmonic.txt
      time.txt
  documentation/
    IntroductionToCompLab.odp
```

Cloud Storage

- The lab book and key results are stored in Google Drive: **https://docs.google.com/document/d/1IYD...**

Repository Storage

A GitHub repository stores:

- Source code
- Analysis tools
- Notebooks
- README documentation
- Instructions for reproducibility

This ensures remote access and versioned backups.

Access is restricted to the researcher unless open-sourcing is required.

3.3 Backup and Preservation

To ensure data integrity:

1. **Google Drive automatic backup**
    - All documentation and processed results synchronised daily.
2. **GitHub remote backup**
    - Code and scripts pushed regularly.
    - Version tracking ensures recoverability.
3. **Manual weekly backups**
    - Exported PDF snapshots of:
        - Lab book
        - Final plots
        - Key CSV results

This ensures long-term accessibility even if local files are lost.

3.4 File Organisation and Naming Conventions

All output data is stored **outside** the src/ directory to avoid mixing source code with results.

Directory Structure

```
lab/
  src/
    epic1d.py
    scan_epic1d.py
  results/
    scan_runs/
      csv/
      plots/
      logs/
    single_runs/
      harmonic.txt
```

```
        time.txt
documentation/
    README.md
    IntroductionToCompLab.odp
```

Naming Conventions

- run_Np1000_Nc20_L12.57_repeat3.csv
- omega_vs_Np.png
- gamma_vs_cells.png
- scan_results.csv

Names include **parameters + repeat index** to guarantee clarity and reproducibility.

3.5 Version Control Strategy

Git is used for full version control:

- Each significant development tagged, labelled clearly
  Examples:
  - *v0.1 — initial PIC code*
  - *v0.2 — added diagnostics*
  - *v1.0 — full parameter scan + multiprocessing*
- Commits are atomic and descriptive:
  - "Add noise detection with threshold"
  - "Implement multiprocessing in scan"
  - "Fix path referencing for results"
- Branches:
  - main – stable, reproducible version
  - dev – new features and experimental modifications

This ensures:

- Complete reproducibility
- Protection from accidental overwrites
- Clear project history for assessment

3.6 Data Sharing and Reuse

Since the project is a coursework laboratory:

- Data will be shared only with the module instructor and lab demonstrators.
- The GitHub repository may be made private.
- If requested, simulation data can be packaged and archived.

All code is written to be reusable for future PIC exercises and research.

3.7 Ethical and Security Considerations

There is **no personal, sensitive, or proprietary data**.
No ethical issues apply.
The simulations generate only synthetic plasma data.