**RDM Protocol - Molecular Photonics 15 April 2024**

**Asst. Prof. Dr. R.M Williams**

This protocol guides researchers in managing research data, ensuring data is FAIR (Findable, Accessible, Interoperable, Reusable).

**For Information for BSc, MSc, PhD students and postdocs, see last page.**

**1. Pre-Project Planning**

**Data Management Roles**

Permanent staff members (PI) guide the post-docs, PhD students and BSc or MSc students to do proper RDM, when appropriate. PI’s have the final responsibility that proper RDM is applied.

If funding requires, then the rules of the funder apply, as specified in Data Management Plan of the proposal. In principle all projects that lead to scientific publications or (PhD) thesis need RDM, build around these research outputs (manuscripts, Thesis Chapters). For MSc and BSc project it is decided by the PI, timely before project end.

Use a simple template with essential questions to establish a proper Research Data Management (RDM) workflow. New researchers should consult senior / other group members on their RDM practices.

Schedule a meeting with your local Data Steward within the first 2 months of your (PI) employment.

Choose one of the Data Storage Solutions below:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **FEIOG** |  | **ICTS** |  | **Office 365** |
| **Name** | Institute Server | Faculty Storage | Research Drive | SurfDrive | OneDrive |
| **Location** | UvA | UvA | Cloud | Cloud | Cloud |
| **Size** | on request | on request | 1-10TB | 500GB | 1TB |
| **Backup mechanism** | None (unless set by institute)? | Snapshots | Versioning | Versioning | Versioning |
| **Confidential** | Yes | Yes | Yes | Yes | No |
| **Intended Use** | Research | Research | Research | Personal | Personal |
| **Sharing (outside UvA)** | No | No | Yes | Yes | Yes |
| **Access** | CIFS/NFS/S3 | CIFS/NFS/S3 | WebDAV | WebDAV | WebDAV |

In the Molecular Photonics group, FEIOG is use for direct storage of data of all important equipment. OneDrive is recommended for storing of collections of data that belong to one project, before the data is made public via, for instance, FigShare.

Currently, weekly backups are generated from the following lab computers on the FEIOG disc:

Nanosecond Transient Absorption

Confocal microscope

Widefield microscope

TCSPC

Both gasphase lab computers

Biotools VCD

Femto second transient absorption

Perkin Elmer IR computer

Bruker computer IR lab

SPEX computer

Horiba computer

Shimadzu UV/VIS computer

**2. Data Collection and Recording**

**2.1. Data Collection Procedures**

- Follow standardised procedures for data collection to ensure consistency and quality.

- Record metadata, including date, location, and instrument settings, for each dataset.

**2.2. Data Backup**

- Regularly back up data to prevent loss in case of accidents or equipment failure.

- Use version control to track changes and updates to datasets.

- The data measured on our main equipment is directly stored on the Feiogg server for safety and security.

Next to data safety this is also implemented to prevent data manipulation or fraud.

**2.3. Organisation and Documentation**

- Organise data files in a logical and consistent manner.

- Maintain clear and detailed documentation of data collection methods, instruments used, and any changes made during the project.

The primary organization of data that is needed for RDM is focussed on a scientific publication, a BSc Thesis, a MSc thesis or PhD thesis. The organization of data can then proceed according to the order in the manuscript: for every figure the raw and processes or analysed data has to be provided. Figure numbering can be used as a primary organization of the data or one RDM project. This also applies to the Supporting Information of the manuscript.

During a project this can also be organized by means of a weekly electronic lab-journal (in e.g. Word, cumulative is advised) or a monthly report with PPP (Problems, Progress Plans). Template is available from the author.

**3. Data Security and Ethical Considerations**

**3.1. Data Security**

- Protect sensitive and confidential data with encryption and restricted access.

- Comply with data protection regulations and institutional policies.

**3.2. Ethical Compliance**

- Ensure ethical approval for data collection involving human subjects or sensitive information.

- Anonymise and de-identify data when necessary to protect privacy.

**4. Data Analysis and Interpretation**

**4.1. Data Analysis Workflow**

- Document data analysis methods, software, and parameters used.

- Maintain records of intermediate results and data transformations.

**4.2. Quality Control**

- Implement quality control measures to identify and address errors or outliers in data.

- Document any data cleaning or pre-processing steps.

**5. Data Sharing and Collaboration**

**5.1. Collaboration Agreements**

- If collaborating with other researchers or institutions, establish data sharing agreements, including data ownership, access, and publication rights. This can be part of a RDM plan for a proposal for funding.

**5.2. Data Sharing Platform**

- Deposit your research data in a suitable repository or platform, following community standards and best practices.

- Provide metadata and documentation to enhance data discoverability and reusability.

For data sharing within collaborations, OneDrive or SurfDrive can be used.

**5.3. Open Access Considerations**

- Align with the principle of 'as open as possible, as closed as necessary,' as appropriate to your research and subject area.

- Consider embargoes or restricted access when needed.

**6. Post-Project Data Management**

**Archiving Data**

Archive all important research data, including raw and processed data, metadata, and documentation, as outlined in your DMP. (As discussed earlier, this can be first one in OneDrive during the project. This data or a selection thereof can now be made public).

Ensure data are stored in a secure and stable environment for long-term preservation (10 year minimum).

Choose one of the following archiving solutions:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Digital Science Ltd** | **CERN** | **ICTS** |
| **Name** | FigShare | Zenodo | Tape/Data Archive |
| Location | Cloud | CERN Data Centre | Cloud |
| Size | 20GB | 50GB (per dataset) | No limit |
| Stored for | No limit | No limit | No limit |
| Sharing | Yes | Yes | No |
| Cost | Free | Free | On request |

If possible FigShare will be the preferred archiving solution. But, if needed, others can be used as well.

An example of a public RDM data collection can be found here:

<https://www.uva.nl/en/profile/w/i/r.m.williams/r.m.williams.html#anker-publications-datasets>

(go to Publications and than to Datasets if needed)

CCDC, Cambridge Crystallographic Data Centre, is the most established structured RDM type feature in chemistry.

<https://www.ccdc.cam.ac.uk/>

Similar organization means may well be organized within chemistry in the future. PhotoChemCAD, for UV-Vis absorption and emission spectra, is going in this direction.

<https://www.photochemcad.com/databases/common-compounds>

This is example of figshare data:

<https://uvaauas.figshare.com/authors/R_M_Williams/3645226>

Two specific RDM data collections, that are directly linked to a scientific publication are exemplified here:

<https://uvaauas.figshare.com/articles/dataset/Gibbbons-2023/22219864>

This is a RDM data collection that belongs to one PhD Thesis Chapter of D.J. Gibbons. The publication is accepted, and has a DOI.

A second example is a Publication with the POLYTHEA ITN-EJD project from 2022:

<https://uvaauas.figshare.com/articles/dataset/SOCT-ISC-PyrDMA-input/17198408>

This DOI is also specified in the publication:

<https://doi.org/10.3390/molecules27030891>

The whole collection of RDM data that belongs to this HORIZON 2020 project can be found here:

<https://doi.org/10.3030/764837>

**6.2. Metadata Preservation**

- Preserve metadata alongside the data to maintain context and facilitate data discovery.

In general the meta data will come with the output of the equipment. For instance the Origin file of the SPEX data contains all meta-data. Preference is therefore to include the Origin file.

**6.3. Data Citation**

- Assign persistent identifiers (e.g., DOIs) to datasets to enable proper citation in publications.

This is standard in Figshare. See examples above.

**6.4. Data Access Policy**

- Define who can access the archived data and under what conditions (e.g., open access, restricted access).

This can be done with One-Drive or other modes of data storage.

**7. Record Keeping**

**7.1. Maintain Records**

- Keep records of all data management activities, including changes, access requests, and data usage. For Figshare this is part of the system. No extra attention is needed.

**7.2. Document Decisions**

- Document any deviations from the initial DMP and reasons for these changes.

**8. Data Publication**

**8.1. Prepare Data for Publication**

- Ensure that data and associated documentation are ready for publication, including compliance with any embargo periods.

**8.2. Publish Data**

- Publish data in the chosen repository or platform. For many purposes Figshare will be a good choice. There is now ample experience with this platform. Examples can be found as part of the literature tab of the uva homepage of co-workers.

- The data can be cited with a typical DOI. Authors of the related work should be mentioned, but the PI or co-worker will be specified.

**9. Review and Compliance**

**9.1. Periodic Review**

- Periodically review and update data management practices and storage solutions to ensure compliance with evolving policies and standards.

**10. Information for BSc, MSc, PhD students and postdocs**

In modern science you have to store all your important raw, processed and analysis data in an organized (digital) way. This commences right at the start of your project. You are the first and foremost responsible person for the data that you have obtained.

If all goes well, your work will be made public by a scientific publication, or other. With this manuscript you will have to provide a collection of data and processed files. You can ask your supervisor which data is important. In general, they will represent the best spectra that you have been able to obtain. No need to do RDM on lousy data/test runs or failed experiments.

The best spectrum of the day, of the month or so will be the important ones. This raw and processed data will be made available to the whole world. Any-one can have a second look at them, re-interpret or re-measure it.

This will help the advancement of science to go faster than before. Your input is required.