

Supporting the Research Data Management Journey of a Postgraduate Student at the University of St Andrews

Federica Fina

Haley Eckel

Panagiota Spanou

Jackie Proven

University of St Andrews

Abstract

Most research funders have requirements for data management plans and open data to foster good research data management practices. In order to embed these practices in the postgraduate research (PGR) student journey we have introduced the requirement for a data management plan as part of the first-year progress review and the encouragement to make data underpinning theses publicly available. To support students through these processes we provide a suite of training workshops and are available for one-to-one consultations. User feedback and frequently asked questions are used to review and improve our support offering.

This brief report discusses the planning and implementation processes for data management plan requirement and encouragement of underpinning data. It dives deeper into the workflows, especially for the data deposit, and describes training and support available to students. Statistics on training uptake, data management plan submissions and annual trends for data deposit are also presented. The report concludes with lessons learnt and the team's plans for the near future.

Submitted 27 May 2024 ~ Accepted 25 June 2024

Correspondence should be addressed to Federica Fina. Email: ff23@st-andrews.ac.uk

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution License, version 4.0. For details please see <https://creativecommons.org/licenses/by/4.0/>



Introduction

Most funders have research data management (RDM) requirements. Funders such as UKRI¹ and European Commission² require submission of data management plans (DMPs) at grant application stage and public availability of research data whenever possible. However, RDM is not limited to funder compliance and open data, but underlies the entire data lifecycle. RDM also includes looking for safe and secure storage solutions, implementing version control, documenting data, and knowing how to handle sensitive and personal data. DMPs foster transparency at all stages of a project, help plan the data lifecycle at an early stage, highlight potential risks for mitigation, and identify limits to data sharing (Barr, Magatova, Denyer, Unsworth, & Smale, 2023).

Postgraduate research (PGR) students are an essential partner in cultural changes required to achieve the full potential of research data management and to implement FAIR principles (Barr, Magatova, Denyer, Unsworth, & Smale, 2023). By training this new generation of researchers we can make good RDM practice the norm amongst researchers of all disciplines.

The University of St Andrews published their first research data management policy in early 2015.³ The RDM service was introduced shortly after that to support researchers in meeting funders' requirements on open data. Pure (Elsevier), our Current Research Information System since 2010, is used as the institutional data catalogue and repository (Fina & Proven, 2017) and a DataCite subscription facilitates registering Digital Object Identifiers (DOI). The RDM service's initial focus was implementing open data guidance for academics on the requirements and deposit of datasets. In 2018, emphasis shifted to the PGR cohort with the 'strong encouragement' of depositing data underpinning theses. In 2022 the institution mandated data management plans for first-year PhD students. Figure 1 shows the current RDM journey of a postgraduate student. In Year 1, students attend a generic induction presented by various University services. For most students, induction is the first point of contact with the RDM team. Assorted RDM-specific sessions and DMP sessions are optional.

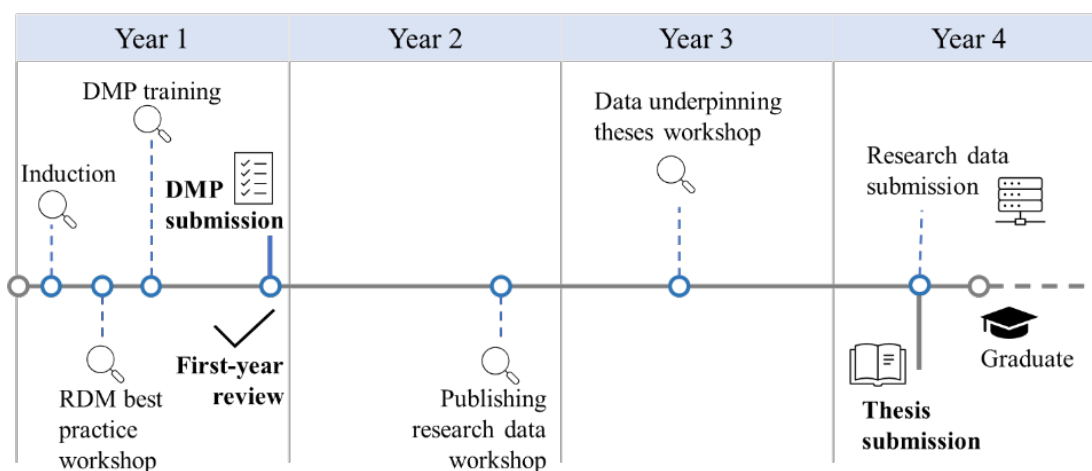


Figure 1 The research data management journey of a postgraduate student.

¹ <https://www.ukri.org/manage-your-award/publishing-your-research-findings/making-your-research-data-open/>

² <https://open-research-europe.ec.europa.eu/for-authors/data-guidelines>

³ <https://www.st-andrews.ac.uk/policy/research-open-research/research-data-management-policy.pdf>

Supporting the Research Data Management Journey

Data Management Plans for First-year Review

A data management plan (DMP) requirement as part of the first-year PhD progress review was introduced for the academic year 2022-2023 after a three-year institutional consultation, approval and implementation process. With the support of the Postgraduate Research Committee, the requirement is now included in the 'Progress reviews and termination of studies for postgraduate research students' policy⁴. The data management plan is one of four mandatory documents that must be included in all first-year submissions. Review submission periods vary from school to school, with the first school submitting in early March. First-year reviews are submitted via the institutional custom-built Module Management System (MMS). As part of the development process, DMP-specific questions for students, supervisors and reviewers were added to MMS. Students are asked to submit a data management plan or explain why it is not required for their research, while supervisors and reviews are asked for their appraisal of submitted plans.

Resources and Support

A DMP template was developed specifically to support students and ensure they consider all the main aspects of data management at the first-year review. The plan includes seven points: 1) types of data, formats and expected volume; 2) collection and transfer; 3) storage, back-up and access; 4) documentation; 5) main risks to data security and mitigations; 6) long-term data retention, preservation strategy and destruction; 7) sharing and publication strategy. Template sections include guidance, with links to further resources. Since students undertaking research involving human participants will need to submit an ethics application form at the same time, the template was developed in consultation with the University Teaching and Research Ethics Committee. The resulting overlap between the template and ethics application form reduces duplication of effort. An assessment checklist was created to guide supervisors through the template, using a checklist of yes/no/not applicable options and highlighting the information they should expect to find in a DMP. Both template and checklist are publicly available on the DMP webpage.⁵

Training sessions are run regularly throughout the year. Early in the DMP cycle, theory workshops cover details of the submission process and teach students what a DMP is and why it is important. These theory workshops are complemented by a training session on research data management best practice. Later in the DMP cycle, writing workshops provide a focused space for DMP work while a member of the RDM team is available to answer questions. Schools can request additional training for first-year students. In addition to group sessions, the RDM team offer one-to-one consultations where DMPs can be discussed prior to submission.

Statistics

In the academic year 2022-2023, 182 students submitted their first-year review. The theory and school workshops were attended by the largest number of students, 124. Writing workshops are limited to 10 participants to allow for a more focused approach. In the first training cycle we run two writing workshops which were attended by 8 students in total. One-to-one consultations were requested by 23 students. Of the 182 students who completed their first-year review, 149 (82%) submitted a DMP file. The most common reason for no file was that the project would not generate data. The majority of all submitting students (77%) discussed their DMP with their supervisors. Supervisors are also expected to answer a few questions; of the 154 (85%) supervisors that completed their reviews only 69% (106) thought that the submitted DMP covered all the aspects specified by the checklist provided by the team.

⁴ <https://www.st-andrews.ac.uk/policy/academic-policies-student-progression-progress-reviews-and-termination-of-studies-for-pgr-students/pgr-progress-review-termination.pdf>

⁵ <https://www.st-andrews.ac.uk/research/support/open-research/research-data-management/requirements-for-postgraduate-students/data-management-plans/>

Submission of Data Underpinning Theses

Process

Since 2018 postgraduate students submitting their theses are strongly encouraged to make data underpinning their work publicly available at the end of their studies, and to include a data access statement in their full text. The data underpinning theses workflows have been developed to align and integrate with the workflows from Student Registry and the e-Theses team. All students are contacted by the RDM team four months prior to their expected PhD end date, and are sent instructions on when and how to start the data submission process. This gives students time to discuss any questions or challenges with the team and to prepare the data for deposit. Students that have made use of data during their project are expected to create a dataset record in the institutional data repository (Pure) as early as possible. Data files should be uploaded only once the examination and the corrections have been completed. Once the data files are uploaded in the research data repository, the RDM team will issue a DOI that the student can include in the full text of their thesis in the form of a data access statement. A DOI is issued only when the student uploads the final datafiles. This is to reduce the risk of students leaving without having uploaded the datafiles.

Students are able to request an embargo for their research data and can choose from one of the following options: 1) embargo on data files only, 2) embargo on data file and description, or 3) embargo on data files, description and title. The RDM team process the datasets accordingly to any embargo requested and redact the metadata if needed. If no embargo is requested, the data files are released immediately. Currently, at the end of the embargo period students are asked by the RDM team whether the embargo on the data should be renewed. To streamline the embargo process and promote open research practices we will implement automatic release of datasets. Students will need to actively contact the RDM team ahead of the embargo end date to request a renewal.

When the students' research involved human participants the RDM team, before accepting the data deposit, need to ensure that the students obtained participants consent to data sharing. Participants are usually informed of the study conditions via a Participants Information Sheet (PIS) and their consent is sought via a consent forms. The students need to provide a copy of these documents to the RDM team for assessment. There have been several instances of students wanting to deposit their data but not being able to do so due to the wording of their consent documents. This was one of the drivers for introducing a first-year data management plan, in the hope that writing a DMP will help students consider future uses for their datasets and solutions for mitigation of risks when participants are involved.

In addition to the guidance sent to the students four months before the end date, training sessions are run through the year. The training sessions cover the submission workflows, the different stages of a dataset record, how to request an embargo and what to consider when the data include sensitive information. The distinction between underpinning and supplementary data is also provided, as this will affect the deposit route. *Underpinning data* are data or digital outputs that are the foundation of a thesis but are not essential to the reader of the thesis to understand it. *Supplementary data* are those data or digital outputs that are an integral part of a thesis and are essential to the reader to understand the document. Underpinning data are treated as described above, while supplementary data are deposited as an appendix alongside the full text on the Theses Research Repository.⁶

Statistics

Between September 2022 and August 2023, 302 students submitted their theses. All students were notified of the encouragement to deposit their research data. Of these, 106 were expected to have a dataset deposited in Pure or elsewhere and 66% (70) of these did deposit their data. Students are expected to add the DOI of their datasets to the full-text of the theses but only about 50% of those with a deposited datasets included a data access statement. There has not

⁶ <https://www.st-andrews.ac.uk/research/support/open-research/research-data-management/requirements-for-postgraduate-students/data-underpinning-theses/>

been a significant increase over the years in the rate of students depositing their research data, but we hope that the DMP for first-year requirement will have a positive affect by raising awareness and helping in addressing some of the limitations of data sharing (e.g. participants' consent).

Lesson Learnt and Next Steps

User feedback indicates students require significant support in understanding both whether a data management plan is applicable and how to represent their projects in one. Projects involving sensitive data, complex projects, and humanities projects are most susceptible to misunderstanding applicability. Across disciplines, safe data storage and data sharing intentions are least understood. Working closely with other units such as Ethics and Data Protection is essential in raising awareness of more challenging aspects of research data management. Unfortunately, supervisors are often unaware of these complexities and require targeted training. Feedback from supervisors indicates that supervisor-targeted training would be helpful, in addition to the assessment checklist already provided on the webpage. Training sessions of first-year DMPs are usually over-booked early in the cycle and booked but under-attended later on. Some students double- or triple-book workshops to ensure a spot but only attend one session. We are currently reviewing booking management strategies to ensure as many students as possible have access to training. The team will also review the submitted DMPs to help inform training and advocacy strategy.

Attendance to training sessions on deposit of data underpinning theses is not high. This is reflected in the questions the team receive at the time of data submission. The team will focus on wider promotion of the training sessions working in collaboration with Schools. The content of the training material will be expanded to cover some of the most common questions students ask and problems they face when depositing, such as providing accurate metadata and organising the files.

Acknowledgements

We would like to thank the MMS team at the University of St Andrews for their help in setting up the questions for the first-year review. Our thanks also go to the e-Theses team who helped us develop the workflow for the submission of data underpinning theses.

References

- Barr, D., Magatova, E., Denyer, G., Unsworth, K., & Smale, N. A. (2023). A Review of the History, Advocacy and Efficacy of Data Management Plans. *International Journal of Digital Curation*, 15(1). doi:10.2218/ijdc.v15i1.525
- Fina, F., & Proven, J. (2017). Using a CRIS to Support Communication of Research: Mapping the Publication Cycle to Deposit Workflows for Data and Publications. *Procedia Computer Science*, 106, 232-238. doi:/10.1016/j.procs.2017.03.020