

Facilitating Re-use of PhD Research Data

Gareth Cole, Jill Evans and Hannah Lloyd-Jones, Open Exeter Project, Library and Research Support, University of Exeter.

The benefits associated with making research data available post-PhD are many. PhD students are increasingly an integral part of research groups; it is frustrating and time-consuming for researchers to search for data on computers and hard drives when the student has left the research group. The cost of reproducing lost or inaccessible data can be high. The student may wish to re-use the data that s/he has produced as part of their PhD in future projects.

Increased visibility of PhD research data can enhance an early career researcher's (ECR) profile as data citation becomes the norm. Career prospects may be facilitated as possible collaborators (including international collaborators and those from different disciplines) can access all the research outputs of the ECR, not just the thesis and/or research papers. New PhD students can face another issue: s/he knows that another PhD student has produced or collected similar data to that that which the new student will use, yet there is no record of this data. This means that data which could have been built upon needs to be reproduced. Discoverable PhD research data also benefits the PhD student's department and institution through association with transparently-produced, visible and citable data.

There are various issues involved in making completed PhD research data available for re-use at higher education institutions such as the University of Exeter. The JISC-funded [Open Exeter project](#)¹ looked at these in our previous paper [Postgraduate Research Data: a New Type of Challenge for Repositories?](#)²

In spring 2012, Open Exeter undertook a Data Asset Framework (DAF) survey to understand the way research data was managed across the University.³ The lack of institutional or departmental policies or guidelines on research data management (RDM) was one of the reasons behind PhD data not being made available. Funding bodies have not strictly monitored compliance with their policies on making data available. Within the University there is not a widespread culture of expecting to make research data available for re-use, although this is commonplace within some disciplines, such as Astrophysics. In spring 2012, there was not a fully working institutional repository for data, nor were there procedures in place for the archiving of PhD research data. Training for PhD students was in pilot stage and research support staff were not confident about giving advice in this area.

This paper will describe how the Open Exeter project has approached these issues over the last year and the successes it has achieved as well as the areas which still need to be addressed.

RDM Policies and Guidelines

¹ <http://as.exeter.ac.uk/library/resources/openaccess/openexeter/>, accessed 4 March 2013.

² <http://hdl.handle.net/10036/4020>, accessed 4 March 2013.

³ See the DAF report at <http://hdl.handle.net/10036/3689>, accessed 4 March 2013.

When a higher education institution has an approved University policy on RDM specifically for PhD students, it is easier to raise awareness amongst PhD students and their supervisors about the significance of data re-use. In fact, the process of putting together a policy, if it is done in consultation with the PhD community, can engage both students and supervisors with the open data movement.

The Open Exeter team developed [University policy on RDM for Postgraduate Research Students](#) (PGRs)⁴ through a [Policy Task and Finish Group](#)⁵, a wide consultation process and approval via the Graduate Faculty Board. All PhD students will need to comply with the policy with regards to RDM from 1st October 2013.

A key element of the policy is that it mandates the registration of selected data produced or collected by the PhD student in the institutional repository. The data itself may be deposited in a subject or national/international repository rather than the institutional repository, or in the case of non-digital data, not deposited at all. The [Open Access and Data Curation Team](#)⁶ will also provide advice to PhD students on analogue research data – see for example, this [case study on lab books](#)⁷ which was undertaken by two PhD students. The policy also states that selected research data should be made available when legally, commercially and ethically appropriate and students can embargo their research data in order to have a period of privileged use of their data. The policy requires that PhD students comply with their funder policy on RDM. PhD supervisors play an important role; they are responsible for providing guidance, conducting an annual research data review with the student, and signing off data for deposit at the end of the PhD (see *Figure 1*).

As well as the institutional-level policy, disciplines and research groups need to develop their own, more detailed, subject-specific RDM guidelines. At Exeter, the Open Exeter team has been working with research groups to help them put together their own RDM policies.⁸ Advice and guidance at this local level increase PhD and supervisor awareness of RDM issues such as making data available for re-use.

User-Friendly Institutional Data Repository

During interviews in the DAF stage of the Open Exeter project and through workshops with the six PhD students who worked on the project it became clear that PhD students need the process of data deposit to be as simple as possible. Exeter's PGR policy does not require data to be deposited specifically in the institutional repository, Open Research Exeter (ORE), which runs on a DSpace platform, but we expect many to deposit here, as there may not be an appropriate discipline-specific repository for their data or they may not have the funds required to pay for archiving in a non-institutional repository.

Exeter's previous repositories have been merged into one, searchable front-end for research data, theses and research papers, which can all be linked together – i.e. PhD data

⁴ <http://hdl.handle.net/10036/4304>, accessed 4 March 2013.

⁵ <http://as.exeter.ac.uk/library/resources/openaccess/openexeter/exeterembeds/policydevelopment/>, accessed 4 March 2013.

⁶ http://as.exeter.ac.uk/divisions/crs/ae/open_access_&_data/, accessed 4 March 2013.

⁷ <http://hdl.handle.net/10036/4393>, accessed 4 March 2013.

⁸ See blog post on Open Exeter's work with the Marine Renewable Energy research group, <http://blogs.exeter.ac.uk/openexeterrdm/blog/2012/06/06/open-exeter-and-marine-renewable-energy-group-policy-case-study/>, accessed 4 March 2013.

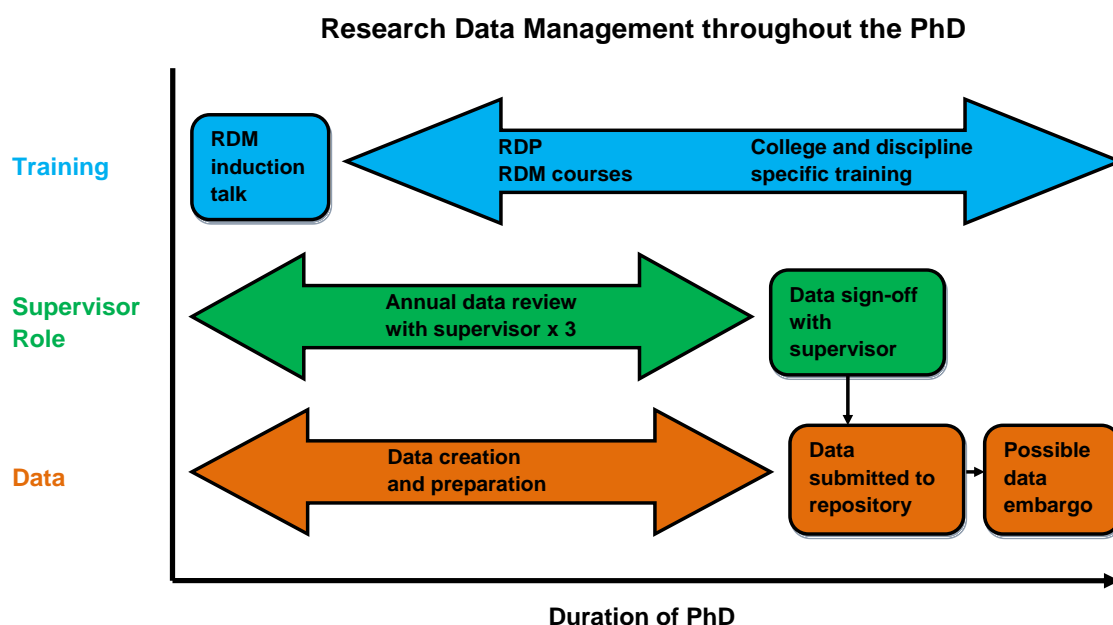
will be able to be found with the thesis that draws conclusions from it. ORE assigns data sets a persistent identifier that can be used for citations and linked to from social media such as Twitter, academia.edu, and so on.

As many PhD students work in multiple locations or off-campus, the deposit solution could not depend on the researcher using University networks or computers. Open Exeter's technical developers have identified the software Globus as a solution to the various problems of depositing research data into ORE.⁹ Globus permits researchers to transfer data from any location to the repository after download of the client software. It also allows secure, seamless single sign-on with University log-in details.

Embedded Training and Advice

In addition to policy and repository development, we needed to raise awareness of RDM across the University's research community and to train PhD students, supervisors and research support staff about re-use of research data. We embedded training courses in existing programmes, such as the [Researcher Development Programme](#),¹⁰ a University-wide programme for PhD students, the [Doctoral Supervision](#) course,¹¹ and College-level PhD inductions and subject-specific training courses. RDM guidance is also available online via the Library pages and is mirrored in the University's Research Toolkit which aims to be a one-stop shop webpage for everything research-related at Exeter.

Figure 1 Research Data Management throughout the PhD.



⁹ See <http://blogs.exeter.ac.uk/openexeterrdm/blog/2013/01/25/dspace-submission-using-globus-and-sword2-update/>, accessed 4 March 2013.

¹⁰ <http://as.exeter.ac.uk/rdp/postgraduateresearchers/>, accessed 4 March 2013.

¹¹ <http://www.exeter.ac.uk/staff/development/courses/coursedetail/?code=20002>, accessed 4 March 2013.

Simple Integrated Procedures

Keeping to our objective of making data deposit simple for the PhD student we worked with the Graduate Administration Office to integrate sign-off of data deposit into the existing procedures that they have for theses deposit. This means that the student, the PhD supervisor and the Graduate Administration Office do not have to deal with additional paperwork. We are currently developing a checklist to help the supervisor at this sign-off point and there is another [checklist](#)¹² available to guide the supervisor in her/his annual research data review with the student. The annual data review can take place as and when convenient for the student and the supervisor. We will add reminder messages about the new RDM policy to the regular messages that the Graduate Administration Office send to students eighteen, twelve and six months before they are due to submit their thesis.

Future issues

There are various issues which have not yet been resolved regarding the sustainability of using the institutional repository to store PhD research data. It is not normally possible to build the costs of data archiving into grant applications for PhD students, although when they are part of a larger research project, this may be allowed. We can estimate how much data will be deposited by PhD students, but until the policy is implemented it is difficult to know how quickly the repository will fill up and how quickly more storage space will be needed.

In addition to the costs involved in data storage, facilitating the re-use of PhD research data requires staff to train, update policy, engage researchers and to oversee the curation of research data and the technical aspects of the repository. At the moment we are awaiting the outcome of proposals for permanent posts which will be responsible for these tasks. If the proposals are not accepted, the level of service and training will be lower than what is desired.

Conclusion

Open Exeter has prepared the ground for widespread re-use of PhD research data at the University of Exeter. University policy, a simple upload solution to the institutional repository and clear, integrated procedures facilitate the process of making research data openly available online. Embedding training and awareness-raising in established training programmes with PhD students, supervisors and research support staff has meant that students who started in the 2012 academic year should be aware of the importance of preparing their data for re-use (e.g. documenting and storing their data according to good practice) from an early stage of data collection or creation.

However, the way in which the data repository will be financed in the longer-term is unknown and a permanent service which promotes the re-use of PhD research data will depend on future staffing levels. Above all, it remains to be seen how the procedures planned will work when the PhD policy is implemented in October 2013.

¹² <http://hdl.handle.net/10036/4214>, accessed 4 March 2013.