



JISC

OPEN EXETER PROJECT

REPORT ON THE PGR AUDIT

Author: Open Exeter Project Team
Creation Date: 11/02/13
Last Updated: 30/09/13
Version: Final

Contents

Introduction:	3
Completing the audit:.....	3
Analysis methodology:.....	4
General findings:	4
Analysis by question:	5
<i>Have you amended any data files this week which you had previously created? Did you need to keep the original version as well as the new version? Please give details.....</i>	5
<i>What file formats are the data you created this week? Please state both electronic and paper.</i>	6
<i>What is the approximate total size of the files you have created this week?.....</i>	6
<i>Approximately how many files have you created this week?</i>	7
<i>Approximately how many files that you created previously have you changed this week? .7</i>	7
<i>What equipment did you use to create your data this week?.....</i>	7
<i>Where was this week's data created? I.e., home, office, field trip, etc.</i>	7
<i>Is any of the data you created this week confidential or sensitive?</i>	8
<i>Have you encrypted any of your data this week?</i>	8
<i>Have you had to transport any of your data this week? E.g. From field work to office. Please give details.</i>	8
<i>Where is the data you created this week saved/stored? Please give details.</i>	8
<i>Has the data you created this week been backed up? Please state where and how?</i>	8
<i>For how long will you be using the data you created this week?.....</i>	9
<i>Does any of the data you created this week need to be shared? Please give details.</i>	9
<i>Is any of the data you created this week likely to be needed to be accessible once your current project has ended? Please give details.....</i>	9
<i>Have you had to use anyone else's data this week? Please give details.</i>	10
<i>Please give any other information on Research Data Management you think will be of use.</i>	10
Unexpected results from the audit:.....	12
How the audit could be improved:	13
Conclusions:.....	13

Open Exeter Report on PGR Audit

Introduction:

As part of the [Follow the Data](#) strand of the [Open Exeter Project](#) the project's PGR students were asked to complete an audit every week on different elements of Research Data Management (RDM). Each week the PGRs were asked to answer the following 17 questions:

1. Have you amended any data files this week which you had previously created? Did you need to keep the original version as well as the new version? Please give details.
2. What file formats are the data you created this week? Please state both electronic and paper.
3. What is the approximate total size of the files you have created this week?
4. Approximately how many files have you created this week?
5. Approximately how many files that you created previously have you changed this week?
6. What equipment did you use to create your data this week?
7. Where was this week's data created? I.e. Home, office, field trip etc.
8. Is any of the data you created this week confidential or sensitive?
9. Have you encrypted any of your data this week?
10. Have you had to transport any of your data this week? E.g. From field work to office. Please give details.
11. Where is the data you created this week saved/stored? Please give details.
12. Has the data you created this week been backed up? Please state where and how?
13. For how long will you be using the data you created this week?
14. Does any of the data you created this week need to be shared? Please give details.
15. Is any of the data you created this week likely to be needed to be accessible once your current project has ended? Please give details.
16. Have you had to use anyone else's data this week? Please give details.
17. Please give any other information on Research Data Management you think will be of use.

The audit began before the [DAF Survey](#) had started and its results analysed. As such, the project team had to make educated assumptions on the questions that they wanted the PGRs to answer. The questions were designed to cover all aspects of data creation, storage and sharing. In addition, as can be seen, question 17 was a catch-all to cover anything else that the PGRs thought might be relevant.

Initially it was expected that the audits would continue for 12 months. However, for reasons which will be described in the report it was decided that enough information had been collected after six months and the audit collection was stopped. As a result, the audits ran from January 2012 until the start of July 2012.

Completing the audit:

Each PGR was asked to complete the audit once a week. At the start of the audit the PGRs were sent an Excel spreadsheet template which had the questions in column one and were initially asked to upload the audit to a [Dropbox](#) folder at the end of every week. However, after a couple of months the PGRs were asked to email the audits each week to the Data Curation Officer because of ongoing network issues with Dropbox (although a couple of the

PGRs continued to use Dropbox to send the audits as well as via email). The forms were then transferred by the Data Curation Officer to the Library's [N:Drive](#) to ensure that they were securely stored and backed up. An Excel file was created for each student to which each week's audit was added. The original files were kept as well as the combined audit form.

There were also fortnightly meetings between the Data Curation Officer and each of students. These meetings allowed issues raised in the audits to be discussed in more detail.

Due to work commitments, conferences, illness, holidays etc. not all the PGRs were able to complete an audit each week. As such, it is not possible to directly compare each student's work. However, this was not the aim of the audit which was an attempt to uncover how PGR students at the University of Exeter worked and if any common practices could be discerned or if there were discipline specific factors that affected data management.

In addition, one of the PGR students working with the project had to drop out so their audits have not been included in this analysis because they were only able to complete a couple of months' worth of audits.

Analysis methodology:

The analysis and first draft of this report were written by the Data Curation Officer. Due to other project commitments the results of the audit could not be fully analysed as soon as the audits were completed. The initial findings and the audit template were posted in a post on the project's blog in [August 2012](#).

Analysis was helped by transferring each of the combined spreadsheets into a single spreadsheet which was then imported into an Access Database. This enabled each question to be easily analysed for the answers of the six PGRs. The reports that were subsequently created within the Database were then exported back into Excel. However, due to migration issues it was necessary, at times, to refer back to the original answers in the Excel spreadsheets.

General findings:

Each question will be analysed in turn in the next section. However, some general points can be made. A number of these have been stated in the August blog post but bear repeating for consistency:

1. All of the PGRs worked in phases. Thus, if were in a data collection phase they did very little background reading and vice versa.
2. Data was created and stored in multiple places by each student. This is examined in more detail below in the analysis of each individual question. However, it does support the findings of the DAF survey as regards the diversity of data collection and storage.
3. Similar issues, e.g. large file sizes and storage, are faced by students of all disciplines. Another common issue is a large number of small files created and resultant questions on how to organise these. A third commonality is versioning.
4. If Microsoft Office applications are excluded, comparatively few file formats were used by the PGRs. The most formats used in any one week was eight (including Office applications) and this was a rare occurrence.
5. All of the students stated at some point in the audit that they have learnt something from working with the Open Exeter project.

6. The audit is a useful exercise for both identifying trends and also making students/researchers think about their research data management practices. By having to write down what they did on a week by week basis it raised awareness of some of their practices that they would otherwise not have thought about (see below for more specifics).
7. The audit could be used as a two way process with the PGR students asking questions and raising issues. This occurred more than the project team was expecting but it meant that live issues could be investigated and real examples could be used rather than abstract exemplars when creating training and guidance.

Analysis by question:

Have you amended any data files this week which you had previously created? Did you need to keep the original version as well as the new version? Please give details.

This question provoked some of the longest answers from the students and also demonstrated that by asking these questions of them, the students began to realise how they could change and improve their working habits. For example, in the first week of the audit one student wrote,

"I kept the original version and the new final version. I initially kept the original version in case I needed to refer back to it. However, now I realise I must have done this process lots before and not gone back and deleted the original when I'm done with the project thus taking up more precious space on my harddrive!"

By half way through the audit, the same student was writing "Yes I amended some word documents. I deleted some of the originals and kept the new versions." The change in working practice is clear. The importance of versioning was also stated by the same student when they update some files: "I kept the originals and also saved updated versions as they are intended for different audiences so it is uesful [sic] to hold on to two separate versions."

Also in week one, another student was aware of versioning issues and had previously identified their own solution, "I have reworked my research notes on docs.google. I use this website in order to keep track of changes while avoiding file duplication (since I am terrible at that). I also find it useful because it saves automatically."

A third student has also previously identified what may be termed good working practices. In week two of the audit they wrote,

"I have kept the original and label them with revision numbers _R0, R1 etc. I keep them just incase [sic] i [sic] realise at a later stage that i [sic] have made a mistake so that i [sic] can refer back to a previously "working version". These originals do not need to be kept on a "live" system and can be archived and removed at a later stage."

During the course of the audit, this student also worked on some contract work (outside of their PhD studies). They wrote:

"Interesting working with a team of people with no official version record. A DMP would have been useful for this process. Two out of three used dropbox, but only

for sharing updated files rather than keeping updated version in one place due to lack of space in one of the accounts.”

Another student also had version control in place early in the audit:

“Yes I have amended 5 different original data sheets in Microsoft Excel this week. I have recently collected data for my first PhD study. I had to add this data to the data I already had before Christmas. The data sheet I am collating the data in has formulas built in. For this reason I keep the original every time I update it encase there is an error when saving the file. I keep the original in a separate folder on my laptop under "Original Documents".”

In one-to-one talks with this student it became apparent that the version control was led by the group leader who encouraged their students in best practice.

The difference in working practices between students can also be seen in an answer from a fifth student (this answer came roughly halfway through the audit):

“I amended an ethical review proposal. It probably would have been good for me to keep version 1, version 2 etc so I could see how they evolved but I don't tend to bother doing that. I tend to have a first draft which I think is passable but I'm not particularly satisfied with and then when I redraft it, I just make changes to that file rather than create a version 2. I suppose there must be a record of the original somewhere though e.g. in an e-mail. Perhaps as part of research data management training it would be worth highlighting the utility of keeping multiple versions of the same file so that you can look back at the changes. It's not somethng [sic] I really think about very much.”

Interestingly, although this student was aware of best practice, i.e. version control, they did not use it personally.

What file formats are the data you created this week? Please state both electronic and paper.

All of the students used Microsoft Word and Excel (either in 2003 form or 2007 i.e. .doc and .docx). Three of the students used what might be termed non-standard software and file formats. Thus, one used ArcGIS to create images. Although these were saved as JPEGs so would be accessible to other users, the raw data files could only be opened on ArcGIS. Another student, who uses specialised lab equipment, created files in a .file format. The same student also used SPSS and GraphPad as software which also create proprietary formats. The third student created files associated with their C++ programme. However, in their case the results “are computed as .txt and .inp for raw data and .inp and excel and word for processed work” (.inp is “essentially notepad .txt files”).

Other file formats used included .pdf, .ppt, .tiff, .mov, .wav, online blogs (Blogger in the case of our student) and Nvivo.

Half of the group also used hardcopy (paper) at some point during the six months of the audit. These included annotated notes on articles that had either been photocopied or printed out, questionnaires, consent forms, and “data sheets”.

What is the approximate total size of the files you have created this week?

None of the students created what may be termed big data. The largest amount of data created in a week from all the PGRs was 1GB. However, one of the students, who creates

sound and video files, did create data of 550MB, 450MB, 600MB, 700MB, 600MB, 550MB, 700MB, 750MB, 650MB and 770MB in ten consecutive weeks. The same student stated that they needed online solutions or a second external hard drive to back up their data (see below). This totals over 6GB of data created in 10 weeks.

Another student raised a useful point in their answers to this question. They created c0.5GBs of data per week but they actually only kept between 20MB and 100MB because the programme they were using automatically overwrote the data that wasn't needed.

Approximately how many files have you created this week?

The number of files created by the students in a week ranged from 0 to 3724. The latter high number is explained by the fact that these files were created by a computer programme and the student explained a few weeks earlier that the programme had automatically "over-written 80% of them". However, this is still a substantial number of files to store each week (although each file was fairly small). Due to the answers on this question we are able to get an idea of the total number of files that may be created by PhD students. Thus, the three first year students created 542, 361 and 122 files over the course of the audit. In contrast, one of the third year students only created 32 new files over the six month period (and 10 of these were in a single week).

Approximately how many files that you created previously have you changed this week?

The answers to this question are also useful in understanding how PGR students work. The same third year student who created 32 new files edited 38 files that had previously been created (although as they were writing up these could have been the same files each week). The first year students above edited 121, 205 and 96 respectively. These numbers show that file organisation and version control are important. This has also been mentioned by the students in the fortnightly meetings.

One of the students wrote that they had amended "all of them [files], although I have kept previous revisions for security". In other weeks the same student only amended two or three files. This was the student whose programme automatically overwrote files.

What equipment did you use to create your data this week?

Generally the students created their data using the same equipment week by week. Thus, if they used a PC in week one they were likely to use a PC in weeks two, three and four as well. The most pieces of equipment used by any of the students in a single week was six.

Three of the students used what may be termed non-standard equipment (specialist equipment needed for their research) as part of their research. Two of them used cameras/video recorders and one used specific lab based equipment.

Where was this week's data created? I.e., home, office, field trip, etc.

All of the students created data in more than one place i.e. home, office, lab etc. This supports and reinforces the findings of the DAF survey. The one student that created data in a lab analysed the data in either their office or at home. Half of the students also created and analysed data whilst abroad (in France, Canada and the USA). One of the students (in the College of Humanities) did not refer to an "office" when asked this question. When on campus they either worked in a "PC room" or "postgrad [sic] study carrel". This supports the wider findings of the Open Exeter project that not all PGRs have dedicated PCs or even office space, particularly in the College of Humanities. The Open Exeter Project has also

found that working in multiple locations can create Data Management issues such as version control, back up and synchronisation difficulties. These difficulties were faced by the students in the audit but, by using applications such as Dropbox, the students were able to alleviate the issues.

Is any of the data you created this week confidential or sensitive?

This question provided some of the shortest answers but also divided the six students the most. Two of the students simply answered “no” every week to this question. Two of the students answered in the positive every week to this question (one of whom expanded by stating they had both paper and electronic data which was sensitive or confidential). One student answered no for every week except for three when they answered “Don’t know”. The sixth student mainly answered no but did mention that they had created a spreadsheet of student marks from their teaching commitments.

Have you encrypted any of your data this week?

Four of the students did not use any encryption at all. One stated that they did for some electronic files (mainly contact details and code documents) but did not “have a secure location for paper documents”. This latter comment reinforces the findings of the DAF survey where a number of respondents stated that they did not have secure storage for their paper records. The final student did not use encryption at the start of the audit but by the end of it they were using encryption following discussion with both the Open Exeter team and [Exeter IT](#).

Have you had to transport any of your data this week? E.g. From field work to office. Please give details.

Five of the students involved in the audit stated that they transported data at some point during the six months of the audit. In some cases this involved moving data from a camera or video recorder to an external hard drive. In other cases it involved moving data between a lab and an office or an office and the student’s home. One student stated that they did not have to transport any data because, “all data goes into dropbox which is updated on all computers.” However, they did “have to be ruthless when I get to my 4GB limit and move some superceded [sic] data from near finished projects to give more room.” Again, this shows the importance of versioning and data selection. A number of the students also emailed data or files to themselves. The DAF survey found that this was fairly prevalent at Exeter so this result was not unexpected.

Where is the data you created this week saved/stored? Please give details.

Previous work in the Open Exeter Project (particularly the DAF survey) has shown that research data at Exeter was stored in disparate places. The PGR audit adds to this body of evidence. Only one of the students solely had their data stored on the University Server. Other students had the data saved in among other places: Dropbox; External hard drives; privately owned computers (both PC and Mac); usb sticks; and online in its original format i.e. as a blog.

Has the data you created this week been backed up? Please state where and how?

Back up solutions proved as disparate for the PGRs as for the researchers who responded to the DAF survey. One of the PGRs wrote: “The only solution is to invest in online storage or a second external hard drive. However, this can be expensive for students. I’m also looking in to online storgage [sic] solutions.” There was a need for this because the “large

film files...take up too much space on my personal laptop's desktop." Other solutions used by the students included: usb sticks, Dropbox, email attachments, external hard drives; the University U:Drive. These solutions are similar to those used by the respondents to the DAF survey. One of the PGRs used Synch Toy to back up their data from their computer to their usb stick. However, in the third week of the audit they wrote "I have not backed up my data this week. I probably should have but just couldn't be bothered continually syncing my pen drive. I haven't made any important changes." It is hard to quantify how common this type of attitude is but anecdotal evidence from the Open Exeter project would suggest that it is fairly widespread amongst researchers and that formal back up procedures and protocols are uncommon if the data is not stored on the University servers.

For how long will you be using the data you created this week?

In their answers to this question the PGRs gave a variety of answers. One student answered that their data would be needed for approximately two years "although it should be securely stored for longer than that for future reference (20 years)". Another student stated between one week and 24 months – it is not known if the data that was only needed for one week was deleted after that time or if the needs of the project meant that it did need to be kept for longer. One student answered "No" for 11 consecutive weeks. Another representative answer was: "Several weeks until I have either finished analysing them or noticed that there is a bug/problem or update my assumptions or input". Another answered "6 months minimum" for every answer. One student who was writing up their thesis and completing an article for most of the audit stated that the data "hopefully...won't be needed again and the edits for the article were final" in one week and in other weeks gave the deadline for submitting that chapter draft to their supervisor.

Does any of the data you created this week need to be shared? Please give details.

All of the students stated at least once that the data or files they were creating that week were to be shared. Mostly the data or files were to be shared with their supervisors. For example, one student wrote in consecutive weeks that the data was shared, "With my 2 supervisors. I will do this by giving them a memory stick with it on" and "It will need to be shared with my supervisors (3 of them). We usually do this via email if the file is not too large." A second student wrote one week that, "I will need to show my documentary sound experiment to my supervisor. Two methos [sic] of dong [sic] this: 1. Bring my external hard drive on campus 2. Load up the video on Vimeo and put a password on access." A third student wrote, "I may share some of it with my supervisors but will need to filter out some sensitive data." Thus, not even the supervisors of this student will see all of their data.

Is any of the data you created this week likely to be needed to be accessible once your current project has ended? Please give details.

The answers to this question included:

- "Possibly. My programming may be useful to somebody, and the post processed results which will not be published."
- "Yes. My edited film files are an integral part of my PhD project."
- "Yes- All will need to be accessible [sic]. There is a possibility the data will need to be reanalysed in future."
- "I am thinking about storing in a secured place where I can access for 20 years following the end of the project. On the other hand, I do not want to make this sensitive data available to just anyone."
- "It will be needed for completion [sic] of my thesis, and then hopefully re-used for publication."

- “Some of it. Particularly the secondary source material. It would be useful if I could keep that indefinitely.”

Each of the above answers came from a different student. As such, all of them would expect some of their work to be available to them (or others) once their current project is over. The fact that each of the students expected to re-use at least some of their data shows that research data management is important across all disciplines even if the data is not made available more openly.

The use of the data would depend on the accompanying metadata and documentation. In the regular meetings with the student who wrote “My programming may be useful to somebody, and the post processed results which will not be published” it was stated that in order for the programme to be useful for other researchers then the assumptions underlying it will need to be known and made available with the code.

In addition, the student who wrote “There is a possibility the data will need to be reanalysed in future” implies that the raw data will need to be preserved and not just the processed or post-processed data.

Both of these have implications for how and where the data is stored (i.e. who has access) and what should accompany the data in the place where it is stored.

The fact that all the students were expecting to have their data available after the end of their current project has implications for the new [University policies on Open Access and Data Management](#) as well as Funder specific policies. As both Funders and the University have moved to an assumption of openness to research data, it is clear that there needs to be a push to raise awareness of relevant policies to both PGR students and the wider academic community. The Advocacy and Governance Officer has developed a communications plan to publicise the policies and to help ensure that the academic community is aware of both the expectations on them and what they should expect from the University.

Have you had to use anyone else's data this week? Please give details.

Of the six students, five of them used somebody else’s data during the six months of the audit. Only one of the students answered “no” for every week. A number of the students interpreted this question as also asking about journal article reading e.g. “extensive reading of existing research to give contextual background to chapter writing.” However, the same student also stated in a later week: “largely picture images downloaded from the internet [sic], and some scanned images from key texts/articles for illustrative purposes”.

A couple of the students stated how they had acquired the data they were using. This was either by downloading it from the internet (“Audio tracks from online sites and sound libraries”) or being emailed by another student/collaborator (“I have used another PhD's data for a small presentation and to calculate some percentage changes from raw data. I was given this via email”). A third student highlighted one of the current issues with data sharing and why it is important that Universities invest in research data management infrastructures: “The tool is based on other peoples [sic] data from multiple people. There have been issues with inconsistencies etc.”

Please give any other information on Research Data Management you think will be of use.

This question allowed the students to report on any issues they faced that week, what solutions they had uncovered, how they had changed their working practices and to ask

general questions of the Data Curation Officer (both rhetorical and needing a reply – usually provided in the face-to-face meetings). The information contained in these shows how the students changed their working practices over time once they became more aware of certain issues relating to RDM by being involved with the Open Exeter project.

Thus, one student wrote in the first week,

“I have recently put into action some of the suggestions that were made in the initial workshop. I have spent a considerable amount of time creating a robust system of data storage and backup. As I reported in the initial workshop report, I store my secondary source data in folders sub-divided into source type. Within these folders the source material is stored as a PDF file. Previously I had very descriptive (and long) file names which I have begun to convert to the 8.3 convention to ensure they are more transferrable and stable. I think this is particularly important because I have had some corruption in data files from empirical research that I previously created and I don't really know why that happened. I have also tried to ensure that my bibliographic [sic] database is synchronised with my file storage system. This has proved useful when I have come to make use of the data in written work. Finally I have also set up a robust means of backing up my data as I described in relation to question 13. One issue I am currently having is incorporating metadata into my source files. I am currently doing this manually and was unsure whether there would be a more efficient means of doing this. I would also be interested to find out if there is any way of making an easily accessible database of all this source material. These are things I hope to investigate in the coming weeks.”

Another wrote (again in week one), “After speaking with others in my office, no-one knows how to password protect a document or are aware of software such as Dropbox. After explain [sic] what this is to them they are all keen to begin using something like it.”

The students were also able to use this question as a means of passing news and possible advice onto the project team. Thus, in the same week that one student wrote, “I think it would be important for the data storage system to be compatible with all the raw data files. This will allow for other researchers to use this raw data and know that it hasn't [sic] been altered” another wrote, “There is going to be a research technology workshop in SSIS looking at the use of all sorts of technological tools in research e.g. Wikis, Nvivo, Limesurvey etc. There is also a Research Data Management workshop in April in Essex for the social sciences but I'm not sure I'd be able to finance my attendance at that, although it would be interesting.”

The students were also able to record feedback on the face-to-face meetings: “It was interesting talking to Gareth [the Data Curation Officer] at this week's meeting about potential challenges and solutions regarding saving and storing film files.” Another used the question to highlight what to discuss in the meetings: “discuss website, data loss and dropbox [sic] at meeting.”

Technology and the availability (or not) of equipment was also mentioned:

- “Internet connection is invaluable for carrying out research and sharing data which allows PhD researchers more freedom location wise to carry out research. Over the past two weeks, I have been able to carry out the same research in the UK, Ireland and France.”
- “Sending documents via e-mail and skype [sic] has proved extremely useful when building my website.”

- “I have had a meeting with an IT technician regarding the encryption of my school leased laptop, as well as my backup external harddrive.”
- “There are no easily accessible scanners in order to form an electronic copy of paper documents. This would provide a safe back up.”
- “Sites such as dropbox [sic] are great for sharing data when you are working remotely as part of a project. I also found it useful showing documents and experiment findings on my ipad [sic]”
- “I think knowledge about the remote access to our university drive is very valuable to all research student [sic]. If this can be highlighted at the start of a PhD i [sic] think a lot of students would use the function.”

Unexpected results from the audit:

The audit highlighted that the students created and used data that was outside of their core PhD work. The students included data on the audit relating to teaching commitments as well as contracted work. The extra information provided by the students helped to identify the other calls on the students’ time and how their research data fitted in with their other data collection.

In addition to this extra information, the audit also highlighted the pressures on PhD students to present at (and in the case of one student, organise) conferences and to write and publish research articles. In some of the weeks of the audit, a number of the students did not have any PhD work in the audit but had been focussing on these other elements of their workloads.

The additional work also highlighted how even students who do not create sensitive or confidential data as part of their research do create such data when they are teaching. This data takes the form of essay marks (and even the names and contact details of the students involved). It is not known what training or advice students who do teach receive in how to cope with holding data such as this.¹

The data associated with teaching also showed that the students were used to sharing (and/or reusing) other researchers’ material. It also made them aware that their teaching materials (Powerpoint slides etc.) could be available to others in the future. The teaching element also highlighted the use of ELE (the Exeter Learning Environment – a Virtual Learning Environment based on the open-source PHP web application [Moodle](#)) by the students. A number of the students asked questions of the Data Curation Officer concerning ELE which may suggest that training and support is not as clear for this application as it could perhaps be.

The finding of the additional work conducted by PhD students also has an implication when training is provided on areas such as naming and organising files. The students who did do additional work stated that it was important that they kept their PhD work and other work separate.

Although this information was not expected via the audit, it was decided not to ask the students to focus solely on their PhD work because this would not have given such rich results and would also have excluded what are in fact important elements in the life of a PhD student. In addition, all of the students independently of each other included the extra information so they obviously felt it was worth reporting and counted as research data.

¹ Although outside of the scope of Open Exeter the support and advice given to Graduate Teaching Assistants (and other staff who teach) on handling sensitive and private data would be of interest in more general data management.

How the audit could be improved:

As with all tasks, the audit could have been improved. In the final group meeting between the PGR students and the Open Exeter Team, the students themselves highlighted some improvements. One stated that ‘The audit worked well initially, but became a little stagnant at times – when writing for long periods for example.’ Another wrote that ‘I felt the report format was a bit difficult for me (Excel Spreadsheet)’.

The project group also feels that the following improvements could also have been made:

1. The questions could have been changed throughout the audit so the answers didn’t become monotonous i.e. new/additional questions could have been added to the audit.
 - a. However, this would have made the analysis harder and more complicated.
2. There could have been more engagement with the PGRs as they were completing the audit.
 - a. However, a decision was made (as discussed in the section on “Unexpected results from the audit” above) to limit advice on completing the audit so that the students weren’t led into giving a specific type of answer or so that useful information was not left out.
3. The questions may have been phrased differently if we had started the audit later in the project. For example, although “What is the approximate total size of the files you have created this week?” seemed a reasonable question when the audit was launched, it subsequently became apparent that it would be difficult for at least one of the students to answer this question accurately. This student used blogs and other web based solutions to communicate and interact with their participants so the file size was almost meaningless in this context.
 - a. However, there was a need to involve the PGR students whilst they were working with the project (they were only with the project as a group for 12 months and the audit started a month into this period). In addition, if the audit had been tailored for the individual research projects then it would have lost some of its richness in the answers that were provided. Indeed, the inability of a student to successfully answer this question is useful information for the project.
4. Different questions could have been asked in the audit.
 - a. For example, the audit didn’t ask whether the student was using a MAC, PC or Linux based machine for their research data. Thus, compatibility issues were not really covered in the core questions.
 - b. “For how long will you be using the data you created this week?” was not as useful as the project team thought it could be because the students didn’t really know the answer to this question. In addition, as a follow up question asking whether this data had subsequently been deleted was not asked in the audit, the answers the students did give could only ever be approximate and analysed as representative rather than absolute.

Conclusions:

In general, the audit has provided a wealth of information for the Open Exeter Project, especially when combined with the one-to-one meetings held every other week between the students and the Data Curation Officer. By assessing and identifying how junior researchers actually do conduct research it has helped the project to be more focussed in providing for long term training and guidance.

In addition, the audit has contributed, together with other research conducted as part of Open Exeter, to a much fuller understanding of the research environment (and the research community) across the University which would not otherwise have been so comprehensive.

Working with a group of researchers from six different disciplines has also enabled differences in research to be identified. However, arguably more importantly, it has shown that there are numerous commonalities amongst, and between, the students involved. Of course, they are not necessarily representative of the whole of their discipline but, and this should not be under stated, the audit results are the results from actual PGRs working on actual PhD projects. As such, the data is unique in its collection as well as in its richness - no other methodology would have produced the same results.

As with other elements of the Open Exeter project the audit has shown that research data management practices are many and varied across the University. Some of the students already had solutions in place to manage their data but others identified changes they could make in their data management due partly because the audit made them think about some of the issues in detail for the first time. As one student said in week one, "I initially kept the original version in case I needed to refer back to it. However, now I realise I must have done this process lots before and not gone back and deleted the original when I'm done with the project thus taking up more precious space on my harddrive [sic]!" By actually thinking about the issues and being involved with the Open Exeter project, this student has improved the way they manage their data. The challenge now is to expand the education and knowledge base across the whole research community at Exeter.