



JISC

OPEN EXETER PROJECT

SUMMARY FINDINGS OF THE OPEN EXETER DATA ASSET FRAMEWORK SURVEY

Author: Open Exeter Project Team
Creation Date: 30/04/12
Last Updated: 08/08/12
Version: Final, edited for public release

Contents

DAF: Executive Summary & Strategic Recommendations	3
Appendix 1: Summary Findings of the Open Exeter DAF Report.....	7
1. Methodology	7
2. DAF Timeline	8
3. Summary Findings	9
3.1 Participation	9
3.2 Externally Funded Research	13
3.3 Non-Electronic Research Data	17
3.4 Sensitive Data.....	18
3.5 Research Methodologies	20
3.6 Data Types	22
3.7 Data Storage.....	25
3.7.1 Live Data Storage	25
3.7.2 Size of Data.....	27
3.8 Data Back Up.....	29
3.9 Data Archiving	32
3.10 Data Sharing.....	35
3.11 Training.....	38
3.12 Bibliographic Software	45
Appendix 2: Open Exeter DAF Online Survey Questions.....	47
Appendix 3: Open Exeter DAF Interview Schedules and Consent Forms.....	47
Photograph, Filming and Sound Recording Consent	58
Photograph, Filming and Sound Recording: Verbal Consent Form.....	59

Open Exeter Data Asset Framework Survey Findings

This report contains an Executive Summary and Recommendations (4pp), plus a full summary report on Open Exeter's DAF survey (Appendix 1).

DAF: Executive Summary & Strategic Recommendations

1. Context

This report provides the initial findings of the research data management survey and interviews conducted by the Open Exeter project.

2. Method

Adapted from the Data Curation Centre's Data Asset Framework methodology, an online survey was created and follow up interviews were conducted with respondents. The survey was designed to uncover how researchers at the University of Exeter created data, where they stored their data, whether they backed up their data and what happened to their data when the project was finished. As with similar surveys conducted by other HEIs the survey was open to PGR students as well as academic staff.

3. Survey Group

The survey attracted 284 responses and follow up interviews have been conducted with over 50 academics and professional services staff both within the University's central services and individual colleges (IT, Research and Knowledge Transfer, Research Accounting and Finance Teams, College Research Managers and College/departmental Computer Development Officers).

The 284 survey respondents included all types of researcher from PGR students to professors. The respondents and interviewees represent all six academic colleges and included researchers based at all three of the University's campuses in Exeter and Cornwall.

4. Findings

The survey has helped us better understand current research data management practice across different disciplines and within different levels of research experience and the findings will inform the development of policy, governance, skills training and future sustainability of research data management at Exeter.

Given the sensitivities of some of the findings, there will be an internal version of the full DAF report, as well as a public version. All findings are anonymised for reporting purposes.

5. Headlines

As a result of high level engagement with the research community, through Associate Deans and Directors of Research, Exeter's DAF survey attracted a high response rate (relative to other HEIs). This gives confidence on the headline conclusions of this report which include:

- Research data management practice, understanding and awareness varies widely across the institution, though (broadly speaking) understanding is higher in STEM/M and Social Sciences than in the Humanities.
- Few researchers have experience of completing a data management plan.
- Researchers are using a wide variety of storage solutions, with varying degrees of information security measures, and the requirement of many researchers exceeds the standard 20GB network backup space allocation.
- There could only ever be partial cost recovery from grants (via direct or indirect costs) for future staffing and infrastructure for research data management.
- Researchers will require research data management support post-award, underlining the need for thorough, guided data management planning throughout the full lifecycle.
- As funders place greater emphasis on research impact, good data management that facilitates open sharing and reuse will become increasingly important.
- Researchers should be encouraged to cite data in the same way that they currently cite published research; there is evidence that published research that provides access to underlying data is cited more frequently.¹
- There is strong evidence of demand for advocacy, engagement and skills training to raise awareness and competency in research data management (this is important to both researchers and professional services).
- Research data management at Exeter needs to be inclusive of analogue (for example, paper records such as log books) as well as digital data.
- There is significant complexity of data, from formats – including obsolete software and hardware and associated licensing issues – to the size of datasets, to the types and specialist knowledge required to describe and manage the data.

The findings underline that management of research data is an emerging field with high-level complexity in organisational and technical terms. The 2015 timescale for the EPSRC Roadmap for research data management gives a good indication of the time it will take – beyond the completion of Open Exeter in March 2013 – to implement and embed good research data management practice at an institutional as well as an individual level.

The findings also suggest that a lot of data is being held, but not always actively managed. Active management involves forward planning around ethical and legal issues, as well as planned choices about how and when to share data and about retention (how to preserve and for how long?). Responsible disposal of research data is a critical part of active management and essential given storage constraints (on analogue or digital data).

It is inevitable that certain aspects of the research data management lifecycle will fall outside of funding periods with no obvious source of funding to allocate to the associated costs.

¹ Piwowar (2007) <http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0000308>

Given that the results highlight good and bad practice, the project is using the findings to create discipline specific training and advocacy materials. One size does not fit all.

6. Next Steps

Open Exeter is using the DAF findings to inform the next project steps, particularly in regard to planning for sustainability, information, advice and guidance (including skills) and policy and governance. Although the DAF survey was not an audit of research data holdings, it has helped the project team to understand the complexity of the data (analogue and digital) and through the follow-up interviews, a number of data sets have been identified for deposit in the Exeter Data Archive.

7. Strategic Recommendations

The following strategic recommendations arise from the DAF.

• **Infrastructure and Cost Recovery**

1. Appointment of a representative from Exeter IT's Information Governance team to the Steering Group for high level strategic guidance on issues of information security (the FOI/Data Protection Officer is already a member of the Open Exeter Policy & Governance Task & Finish Group).
2. Establish an Open Exeter Sustainability T&F Group to develop fair institutional models for cost recovery and/or future investment in the infrastructure and posts to ensure Exeter's compliance with funder requirements for managing and sharing of research data.
3. Exeter should adopt a holistic approach to research data and develop strategies and solutions that are inclusive of digital and analogue holdings. Strategies and solutions will necessarily include institutional, local (College) and individual responsibilities and commitments.
4. That the Exeter Data Archive will include research data that has been approved for Open Access sharing and re-use (subject to appropriate attribution) and will also include metadata-only records for research data that cannot be openly shared (for ethical, commercial and legal reasons) – where the actual data will be stored as a 'dark archive', managed, but not publicly accessible.
5. The findings of the report are openly shared and discussed with Exeter IT and Colleges and that particular attention is paid to:
 - a. storage needs of live research data.
 - b. responsible sharing and movement between devices of live research data.
 - c. information security issues connected with live research data.

• **Policy & Governance**

6. A University Research Data Management Policy is developed and ratified. This is currently being developed via the Open Exeter Policy & Governance T&F Group. The draft policies (covering researcher and PGR data) have been approved for consultation by the Steering Group.
7. That each research group/project will need its own research data management plan so that local responsibility and ownership is clear. (The

Project is currently piloting this approach with the Marine Renewable Energy research group.)

8. That this is an emerging area of practice and policy and will require post-project focus and resources. This will need to include review mechanisms for the EPSRC Roadmap to 2015 and other funder policies.

- **Skills and Training**

9. That the University provides discipline-specific training on data management issues, with particular priority to data management plans and core data management skills (for example, file management, organising material). That this training is embedded within existing programmes, e.g., the Researcher Development Programme (RDP) and the Postgraduate Certificate in Academic Practice (PCAP).
10. That the needs of PGRs are to be considered in terms of:
 - a. Specific training and support models (RDP, potential for peer to peer mentoring, role of the doctoral supervisor).
 - b. PhD data management plans addressing long-term storage of data (whether this lies with the student or is deposited on the Exeter Data Archive, with the PhD).
11. That training programmes are inclusive of professional services staff in the centre and in Colleges including Library, Research and Knowledge Transfer (RKT), IT, Research and Finance teams (RAFT), Computing Development Officers (CDOs), and so on.

- **Future Staffing**

12. That new responsibilities will need to be accepted into central and College teams to ensure compliance and improved management of the University's research data assets. That future sustainability models for the project are likely to include recommendations for additional dedicated staffing to help manage and monitor institutional research data management policy and practice.

**Open Exeter Project Team
July 2012**

Overleaf: Appendix 1: Summary Findings of the DAF Report

Appendix 1: Summary Findings of the Open Exeter DAF Report

1. Methodology

Following a period of initial research, the Data Asset Framework (DAF) methodology was selected as the most appropriate tool with which to survey University researchers.² The team approach built on findings and lessons learnt by previous DAF implementations at, for example, Edinburgh and Northampton.³

Data collection consisted of two main strands: an online survey, which ran from 14th February until 30th March 2012 and interviews with researchers, administrators and IT support staff. Interviews will continue to be conducted throughout the length of the project, allowing the team to focus on deeper investigation of specific survey findings. In addition, it was felt by the project team that ongoing face to face contact with researchers would aid awareness and engagement within the research community.

The online survey consisted of 34 questions.⁴ The questions asked were tailored for Exeter's needs and the aims and objectives of the Open Exeter project: understanding Research Data Management (RDM) practice rather than counting data assets. The draft survey was piloted with a group of seven Post-Graduate Research (PGR) students working with the project.⁵ By using the PGRs as a pilot group the survey was able to be tested on a group of researchers similar to the target audience. The pilot process proved particularly valuable in identifying which questions needed to be re-worded to clarify understanding and ensure that consistent responses would be obtained.

DAF implementations elsewhere have resulted in limited involvement from the researcher community. In order to encourage survey response, it was decided that a programme of advocacy should be conducted before the survey launch. In the weeks prior to the launch members of the project team met with senior members of the six academic Colleges including College level Associate Deans of Research and Assistant College Managers for Research as well as departmental level Directors of Research. These meetings allowed the team to explain the aims and objectives of the project and to outline the long-term benefits of engagement with project strategy to the College. These face to face meetings allowed the team to build a good rapport with senior College staff and to gain ongoing support for project work.

In addition, the meetings provided valuable insight into the research data issues facing the different Colleges and departments. These insights in turn helped to develop the survey content.

As a consequence of these pre-survey meetings communications that were sent to researchers informing them of the survey, although drafted by members of Open

² <http://www.dcc.ac.uk/resources/repository-audit-and-assessment/data-asset-framework>

³ <http://www.data-audit.eu/users.html>

⁴ See Appendix Two.

⁵ <http://as.exeter.ac.uk/library/resources/openaccess/openexeter/humandata/>

Exeter, were distributed through internal College and department mechanisms. In this way the survey was promoted by a member of the researcher's own academic community rather than an unknown central services staff member. The project team feel that this high-level College involvement greatly helped to raise the survey response rate.

As encouragement to complete the survey a Kindle was offered as a prize for a randomly chosen respondent. Anecdotal evidence suggests that this acted as an additional incentive.

As already noted, follow up interviews will continue for the duration of the project. Up to this point interviews have been semi-structured based on a schedule designed either for PGR students, academic staff, IT support staff or administrative staff.⁶ The interviews are generally of one hour duration (although this does depend on the availability of the individuals involved) and, where possible, are conducted by two members of the project team. Hand written notes are taken by interviewers and interviews are recorded on an MP3 player with the interviewee's consent. Notes are written up after the interview and checked for accuracy by other team members. It was decided early on not to transcribe interviews fully due to the amount of work involved. Both the audio recordings and the interview notes are being added to an NVivo 9 database to aid analysis of the results. Other data collected as part of the Open Exeter project (such as recruitment interviews with PGRs) are also being added to this database.

Participant privacy has been respected at all stages of investigation: survey respondents could opt to remain anonymous; quotations from interviews will be used only with participants' permission; consent is always obtained for recording or filming; a data protection statement was included at the start of the survey.

2. DAF Timeline

It was felt that a launch post-Christmas vacation would be more likely to attract a higher level of response.

Date	Activity
Dec 11-Jan 2012	Creation and iterative evaluation of survey
Beg Feb 2012	Pilot with PGRs
Beg-mid Feb 2012	Final amendments to survey
End Jan-mid Feb	Pre-survey meetings with College managers
14 February 2012	Online survey launched
February 2012 onwards	Interviews conducted with researchers and relevant support staff
30 th March 2012	Online survey closed
3-27 April 2012	Data entry in NVivo
May 2012	Data analysis
May 2012	Report write up started

⁶ See Appendix Three for the schedules.

8 June 2012	Draft circulated for comment
21 June 2012	Summary for Steering Group available
30 July 2012	Public version available

3. Summary Findings

The following sections and charts outline initial results primarily from analysis of the online survey. In some cases, where appropriate, findings have been backed-up by excerpts from follow-up interviews as this information is added to the NVivo database. Interview findings are proving to be a rich and fascinating source of information and the project team aim to make as much of this available as possible over the remainder of the project.

3.1 Participation

The University of Exeter academic structure comprises [six units](#): the College of Engineering, Mathematics and Physical Sciences (CEMPS); the College of Humanities (CHUMS); the College of Life and Environmental Sciences (CLES); the College of Social Sciences and International Studies (CSSIS); the Business School; and the University of Exeter Medical School (the Peninsula College of Medicine and Dentistry (PCMD) at the time of the survey).

The six units are spread over three main campuses: Streatham and St. Luke's in Exeter and Tremough in Cornwall. In addition there are many smaller bases (largely part of the Medical School), for example, the Knowledge Spa at Treliske Hospital, Truro. Responses were received from all sites. The online survey received 284 responses. Responses covered all the academic Colleges and all types of researcher. Results are shown in **Charts 3.1.1 and 3.1.2**:

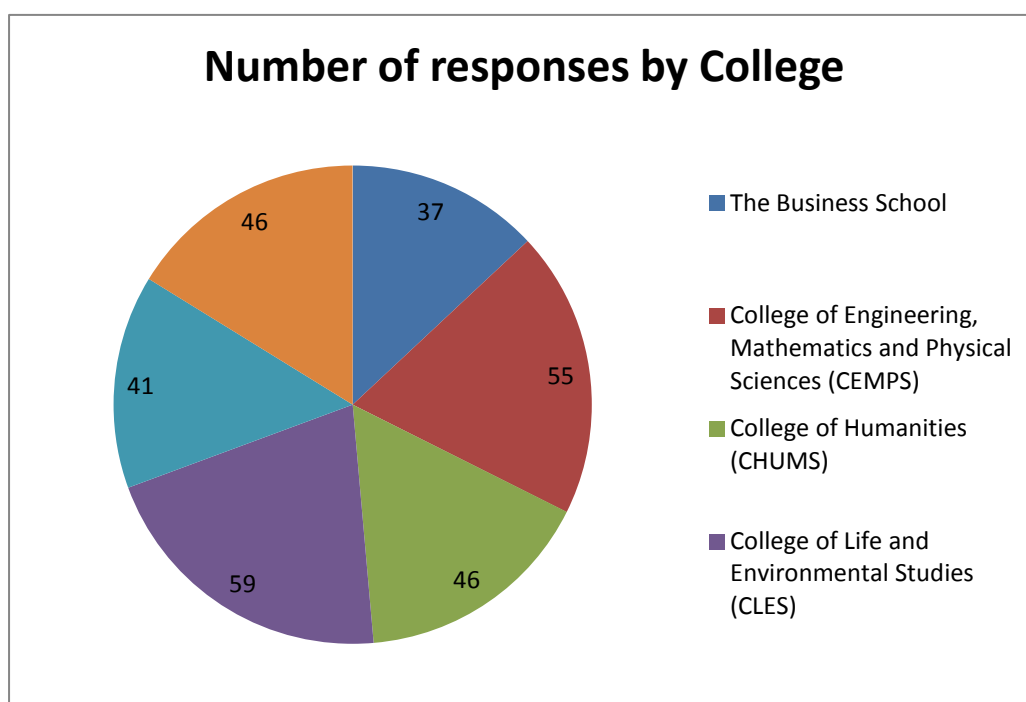


Chart 3.1.1: Number of responses to the survey by College

Initially, the number of respondents from the College of Humanities was disappointing although targeted promotion and advocacy work eventually raised the level of participation. Follow up interviews have suggested that Humanities researchers do not always consider that they use research 'data'. The word is seen as a scientific term and is not as widely used in Humanities as it is in the Sciences. A number of researchers emphasised that language used to describe research is very important, and the way that Science, Technology, Engineering and Mathematics (STEM) and Humanities and Social Sciences (HASS) researchers talk about their work is vastly different. This finding supports the outcomes of previous debates held with PGRs around definitions and meanings of 'data'.⁷

In initial meetings with Business School staff it was suggested that the response rate from researchers in the School could be lower than in other Colleges due to a high staff turnover rate and the fact that previous surveys had attracted low response rates. Although ultimately the Business School did provide the fewest number of respondents, the response rate was not as low as initially feared. However, as **Chart 3.1.7** shows, nearly three quarters of respondents from the Business School were PGR students, far higher than in the other Colleges.

Chart 3.1.2 shows that by far the largest group of survey respondents was PGR students. However, the 46% of respondents fitting this category is broadly in line with comparable surveys.⁸

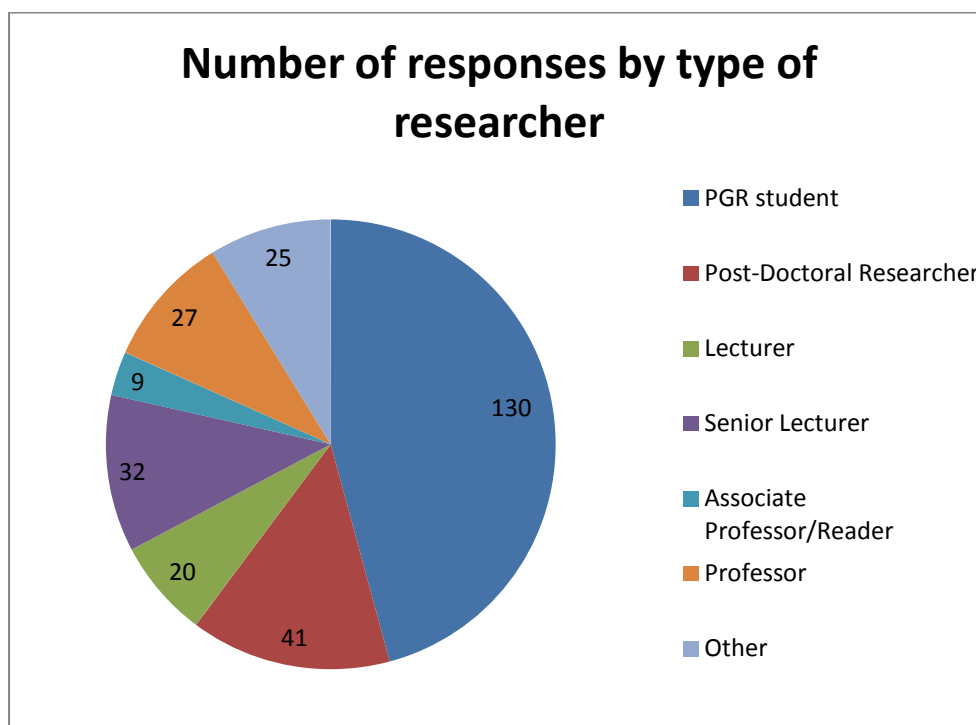


Chart 3.1.2: Number of responses to the survey by position held at the University

⁷ <http://blogs.exeter.ac.uk/openexeterrdm/blog/2011/12/14/first-pgr-workshop/>

⁸ For example, a similar survey at Northampton University in 2010 attracted a response rate of 32.5% from PGR students: <http://nectar.northampton.ac.uk/2736/>

The breakdown of researcher type by College also shows that with the exception of the Business School and the Medical School (PCMD) the percentage of PGR responses were reasonably similar. The results are shown in **Charts 3.1.3 – 3.1.8**.

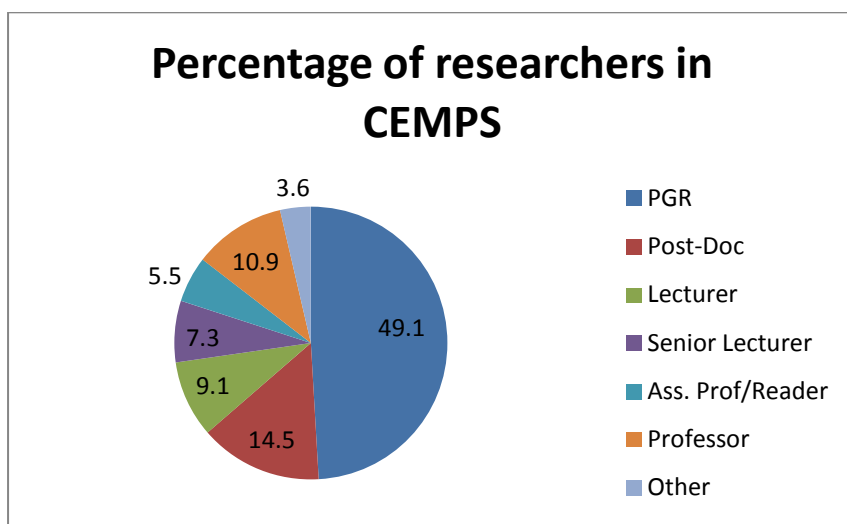


Chart 3.1.3: Responses from CEMPS by position held

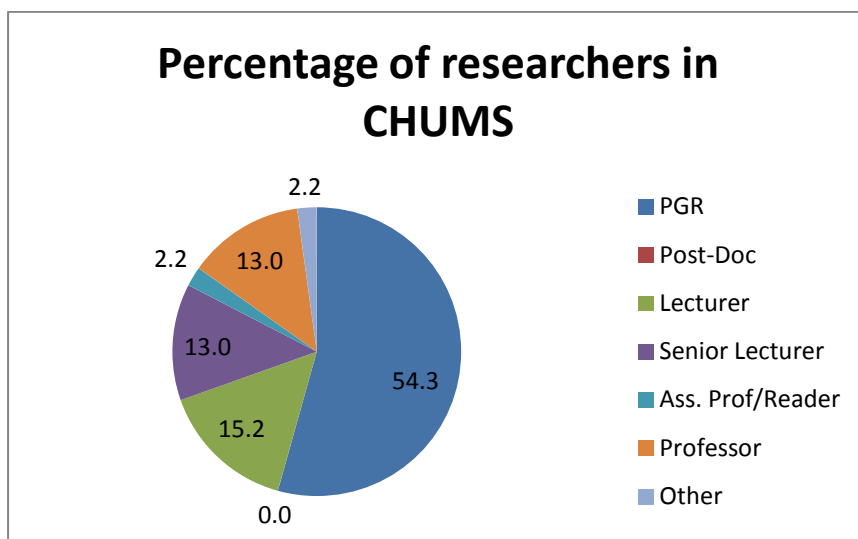


Chart 3.1.4: Responses from CHUMS by position held

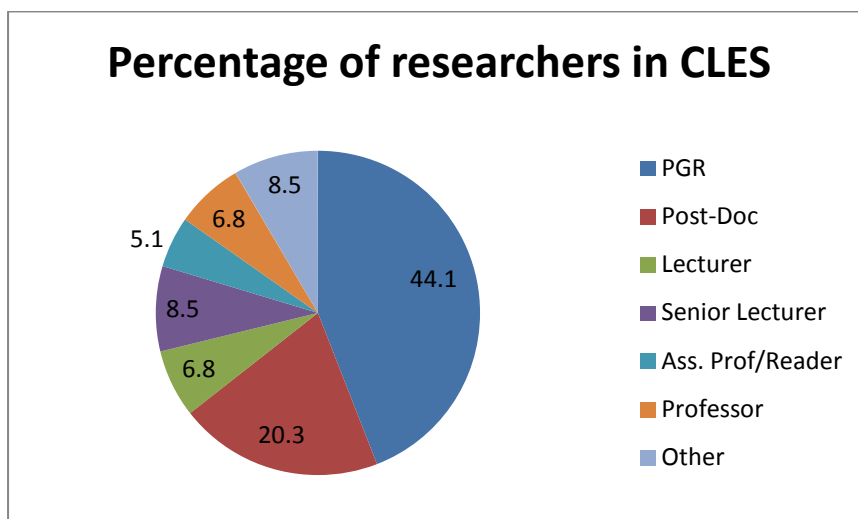


Chart 3.1.5: Responses from CLES by position held

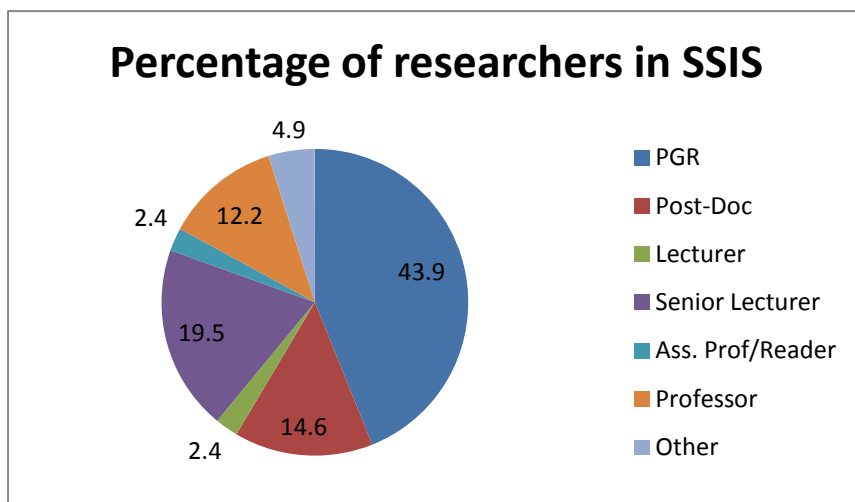


Chart 3.1.6: Responses from SSIS by position held

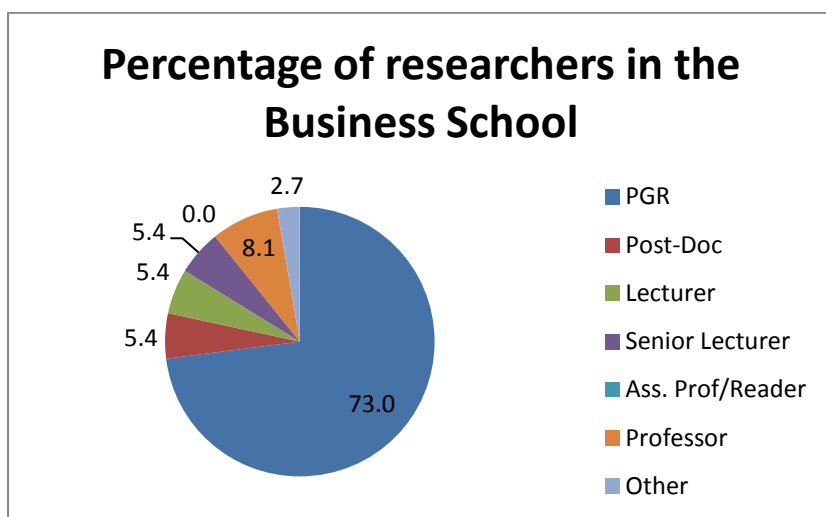


Chart 3.1.7: Responses from the Business School by position held

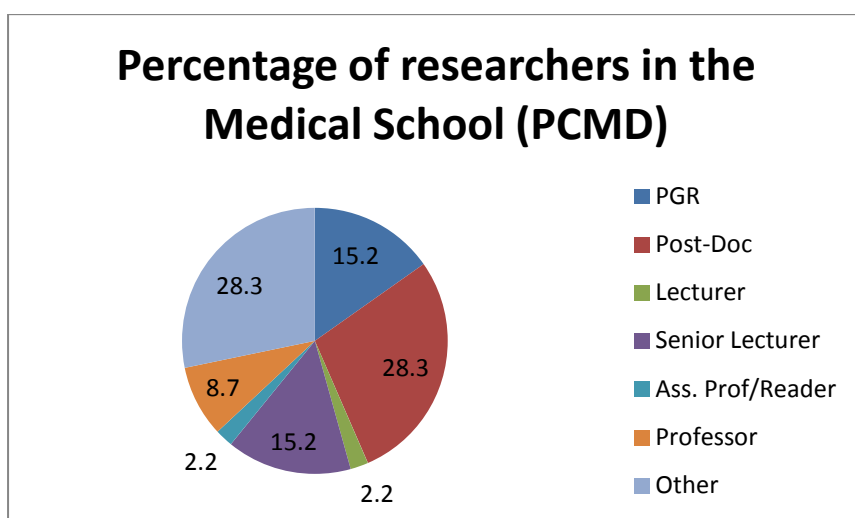


Chart 3.1.8: Responses from PCMD by position held

3.2 Externally Funded Research

The survey asked respondents to categorise how their own research was funded. One of the aims here was to identify the range of external funders of Exeter research as well as Quality-related Research funding (QR) and self-funders (mainly PGR). This data will help to inform priorities for training (subject to different funder requirements) and the debate on future cost recovery. The results demonstrated that an institution-wide approach to research data management is required, irrespective of the different ways research is funded. However, it will be possible to prioritise and segment different audiences for research training and advocacy based on the different funders associated with different disciplines.

Chart 3.2.1 below shows that a large number of researchers considered themselves as not currently being in receipt of externally funded grants (for example, from RCUK) and these include researchers dealing with large amounts of data. This finding emphasises the fact that cost recovery cannot be achieved simply through research grant direct or indirect costs. Data management is also a continuous process, between and beyond grants, and there will always be a need to ensure that research not governed by funder policy complies with good RDM practice. In addition, certain aspects of the RDM lifecycle will fall outside of funding periods but support will still be required.

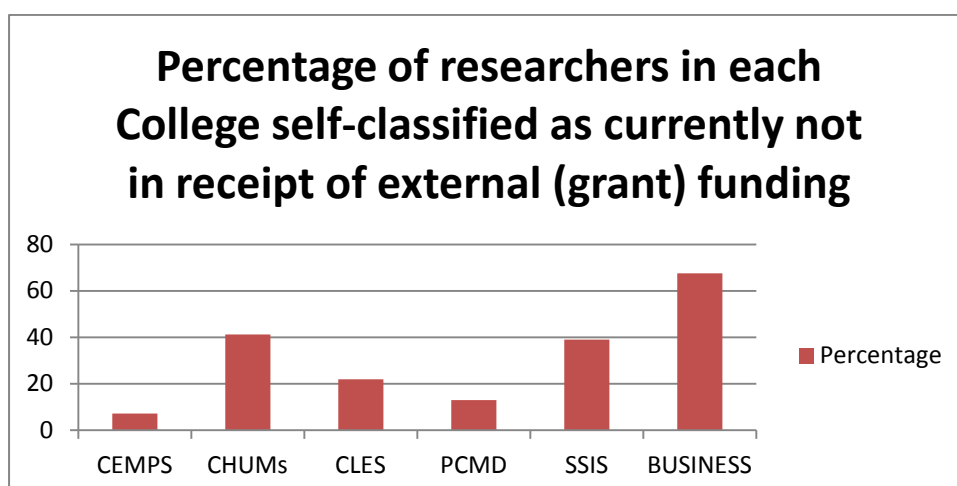


Chart 3.2.1: Percentage of researchers in each College self-classified as not in receipt of external (grant) funding

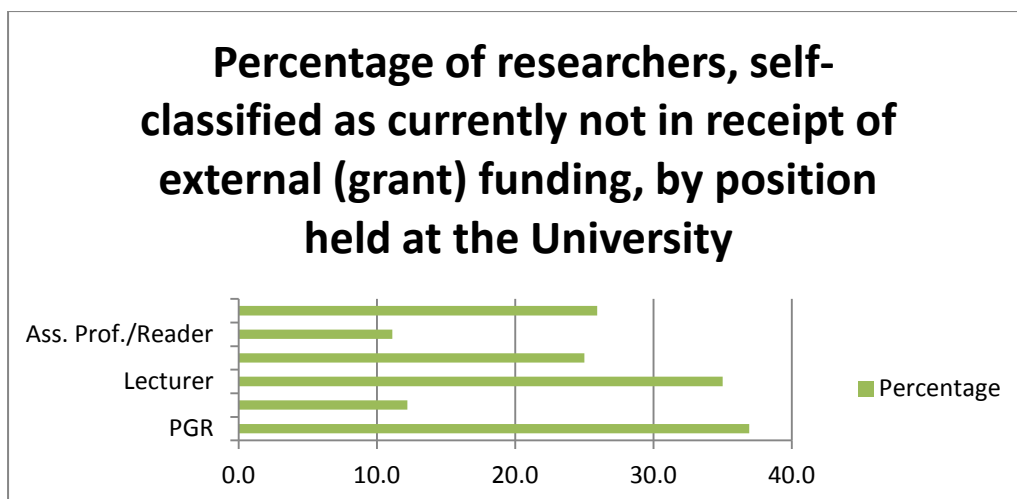


Chart 3.2.2: Percentage of researchers, self-classified as not in receipt of external (grant) funding, by position held at the University

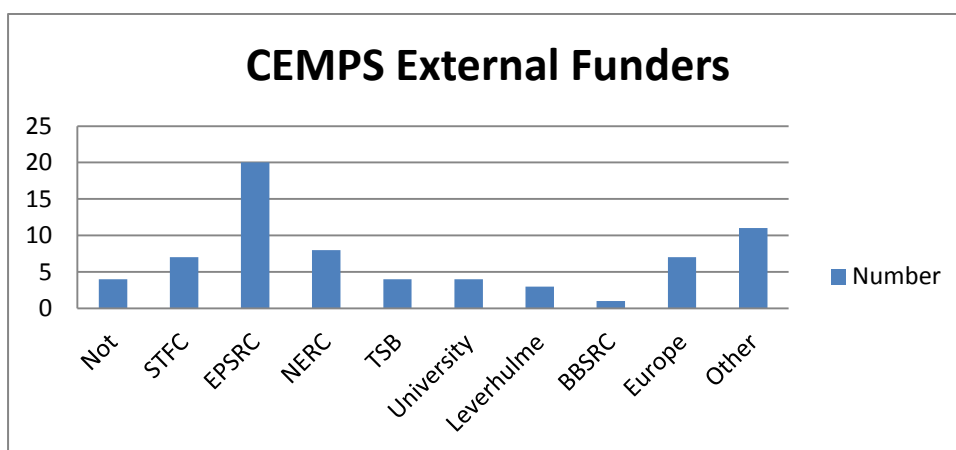


Chart 3.2.3: Funders in CEMPS

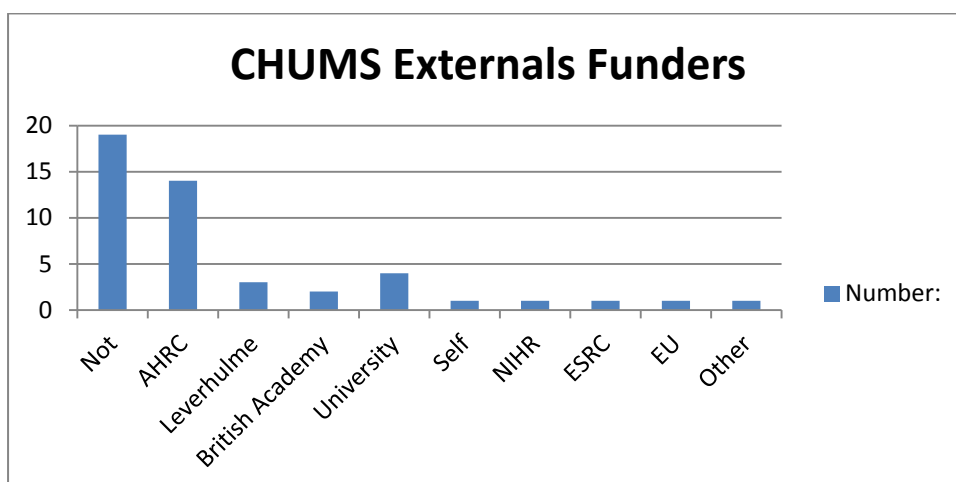


Chart 3.2.4: Funders in CHUMS

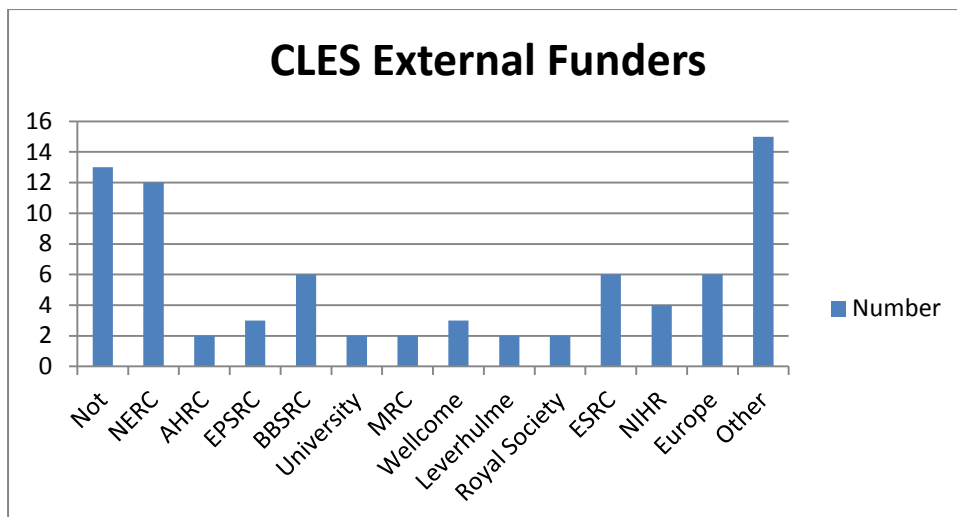


Chart 3.2.5: Funders in CLES

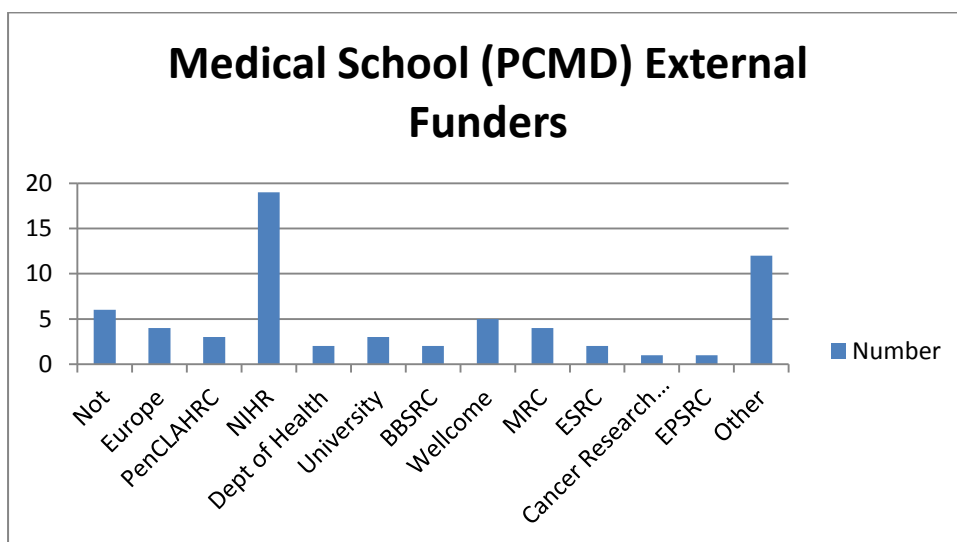


Chart 3.2.6: Funders in Medical School

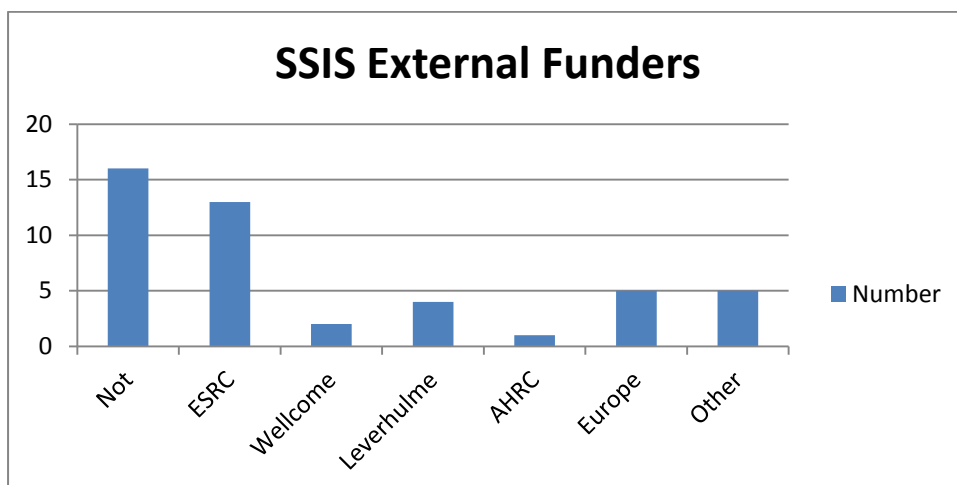


Chart 3.2.7: Funders in SSIS

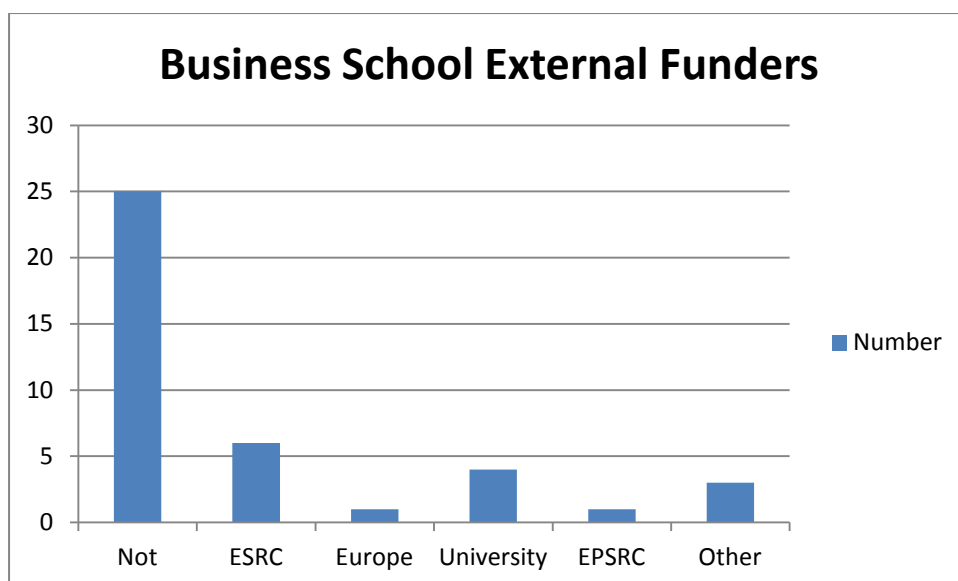


Chart 3.2.8: Funders in the Business School

The reason for mapping Colleges and funders through the survey is to help inform the development of discipline/funder-focused training, advice and guidance. Awareness of major funders' policies in particular subject areas will permit the development of customised training that will ensure compliance and minimise the risk of loss of research income.

The breadth of funders in the STEM/M subjects in particular will make it harder to give generic advice although it may be possible to focus on the larger funding bodies: RCUK; Wellcome; European Union, and so on.

Further analysis is needed in this section to produce a clearer picture, for example, by analysing and treating PGR responses separately.

3.3 Non-Electronic Research Data

Respondents were asked to list the types of non-electronic data that they typically use. A high percentage of respondents, nearly 63%, have data that falls into this category. This includes lab books, field notebooks, interview transcripts, consent forms, “personal research notes”, newspaper clippings and questionnaires. One respondent currently held 6,400 paper surveys and expected to accrue 10,000 by the end of their study. Again, these findings support the outcomes of discussions with the Open Exeter group of PGRs.⁹

These results demonstrate the need for a holistic approach to research data management (and data management plans (DMPs)) which is inclusive of analogue and digital data formats. This is especially important when one considers the expectations of the RCUK funding bodies. For example, the EPSRC states that, “Publicly-funded research data that is not generated in digital format will be stored in a manner to facilitate it being shared in the event of a valid request for access to the data being received (this expectation could be satisfied by implementing a policy to convert and store such data in digital format in a timely manner)”.¹⁰

The Open Exeter project is experimenting with digitisation of sample analogue Science data. Digitisation may be one future option for data sets of the highest public importance and in highest demand. It would, however, be possible for metadata-only records describing and recording analogue research data to be included in the Exeter Data Archive (EDA) to meet RCUK requirements on describing and acknowledging underlying research materials associated with research outputs.

Non-electronic data, or as one Humanities professor phrased it, ‘Old-fashioned notes in folders’, are still an integral element of data collection and data storage. Indeed, **Chart 3.3.2** shows that a substantial percentage of junior researchers, i.e., PGR students, still regularly create and use non-electronic data. This finding indicates that all future policies and storage options should give consideration to both analogue and digital data. As **Charts 3.3.1 and 3.3.2** illustrate, this particular finding is present across all of Exeter’s Colleges and all types of researcher. Responses to the survey suggest that the high percentage of analogue material held in the Medical School is due to legal and ethical requirements to capture and store patient-related information such as ‘patient files’, ‘consent forms’, and ‘questionnaires’. The CLES figure may mainly be due to the extensive presence of ‘lab books’.

The results above underline the fact that data management plans, and training in writing these, need to refer both to digital and analogue data, alongside active data management that includes decisions about retention and disposal. Not all research data – analogue or digital - requires long-term storage and there are considerable savings to be made by educating researchers in effective data evaluation and proper disposal.¹¹

⁹ <http://blogs.exeter.ac.uk/openexeterrdm/blog/2012/01/27/what-is-data-some-responses-from-pgrs/>

¹⁰ Expectation iv: <http://www.epsrc.ac.uk/about/standards/researchdata/Pages/expectations.aspx>

¹¹ <http://www.dcc.ac.uk/resources/how-guides/appraise-select-data#2>

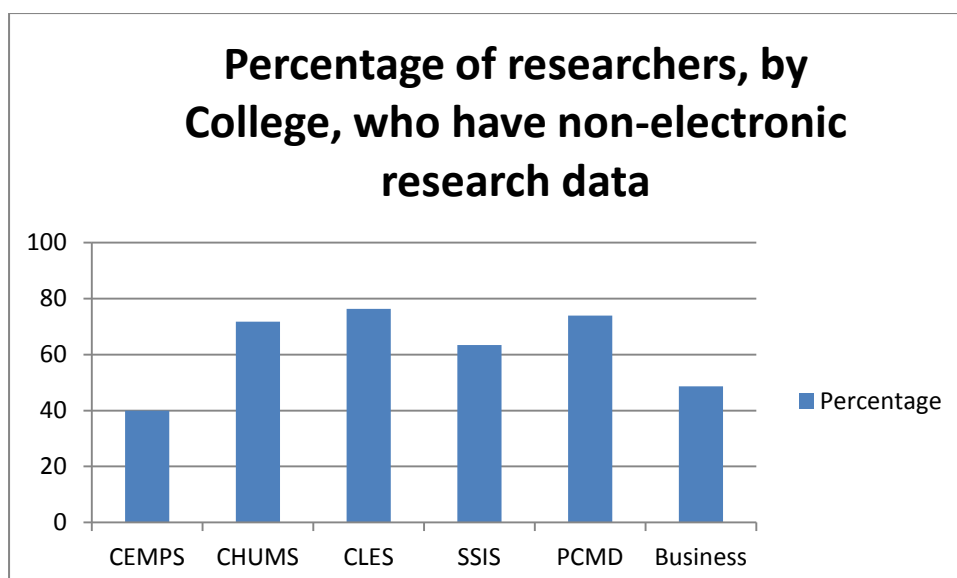


Chart 3.3.1: Percentage of researchers, by College, who have non-electronic research data

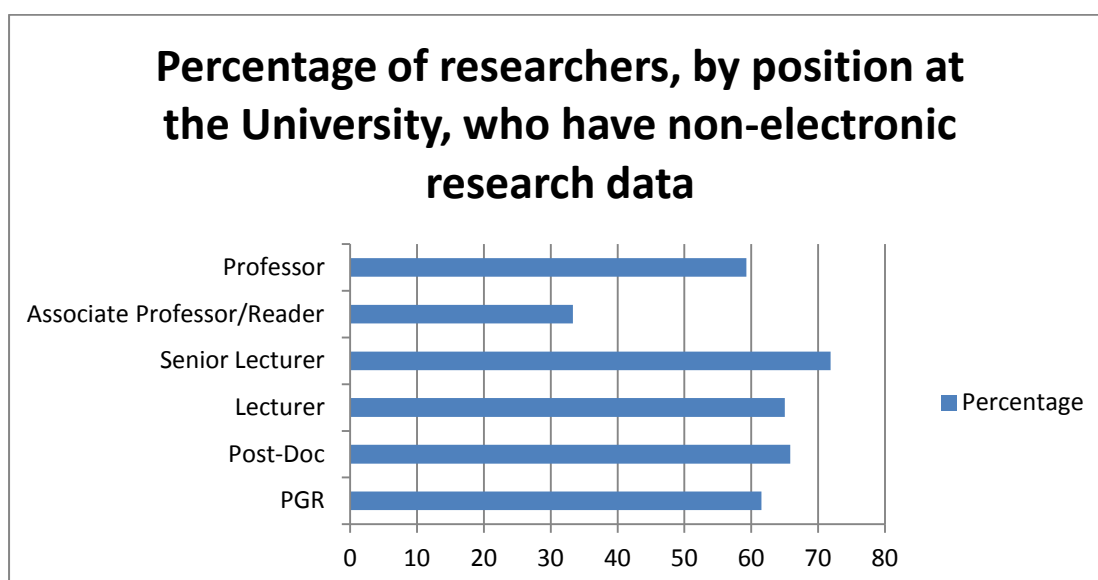


Chart 3.3.2: Percentage of researchers, by position at the University, who have non-electronic research data

3.4 Sensitive Data

Survey results demonstrate that researchers in all Colleges perceive themselves to be creating or using sensitive or confidential data. In retrospect, it would have been useful if the survey had been able to draw out what researchers meant by sensitive or confidential data as there may be a number of different reasons for classifying data as sensitive. In addition, sensitive and confidential data can be very dissimilar in nature and may need to be treated differently. In order to understand the responses to this section more fully, the question is being explored further in interviews with researchers. **Chart 3.4.1** shows that CHUMS contains substantially fewer researchers who use sensitive or confidential data than other colleges, but even here the figure is 20%.

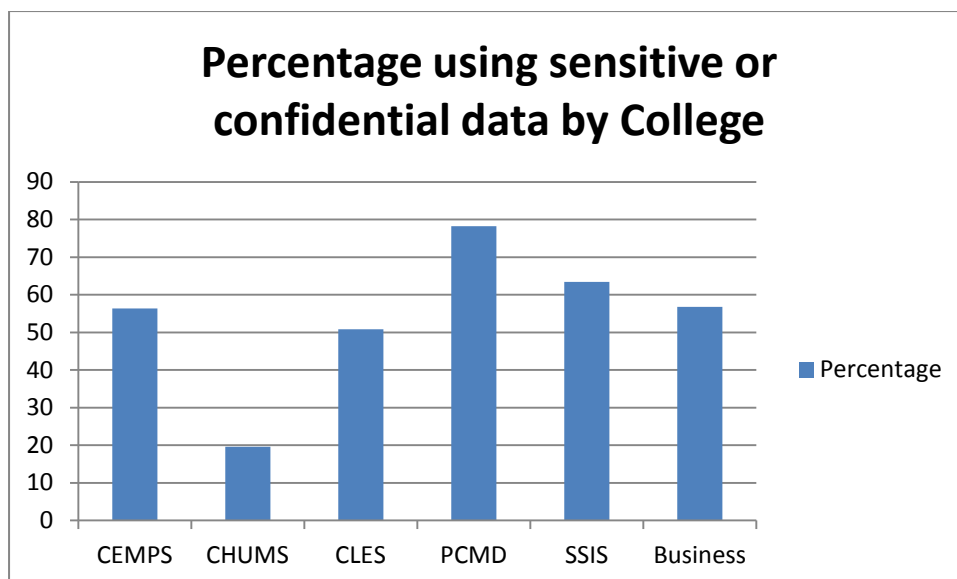


Chart 3.4.1: Percentage of researchers, by College, using sensitive or confidential data

Chart 3.4.2 shows the number of researchers within each college who are aware of their legal obligations to hold data securely. It is clear that the Medical School (PCMD) contains the greatest number of researchers who are under legal obligations and that these respondents are active data managers able to confirm their understanding of the legal and ethical framework for managing confidential data.

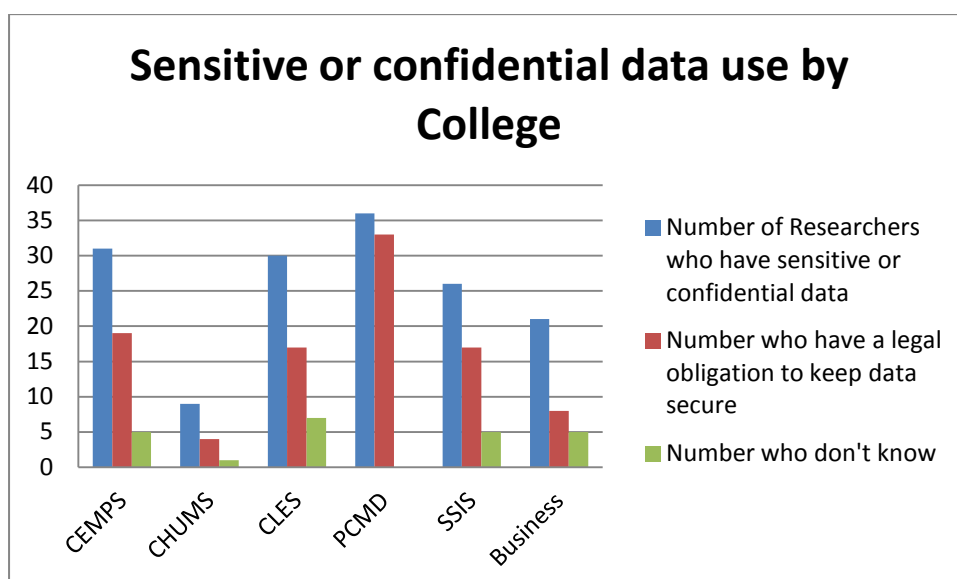


Chart 3.4.2: Sensitive or confidential data use showing how many researchers are under a legal obligation to keep their data secure and how many do not know if they are

38 researchers who produce or use confidential data self-classified themselves as not currently working on external grant funded projects. Of these, 26 reported that they are under legal obligations to keep their data secure. This finding emphasises the complexity of any cost recovery model based on the direct or indirect costs of research grants.

Of the 153 researchers who stated that they hold research data of a 'sensitive or confidential nature' only 39 have data that is encrypted. One of the recommendations of this report might be, through training, to increase knowledge of data encryption options. This could be as simple as encrypting individual files or the whole PC/laptop. As all new University laptops are now encrypted this situation may gradually improve. However, as shall be seen below, a substantial proportion of researchers do not currently use University systems or facilities. One College has considered handing out pre-encrypted USB sticks to new researchers but the cost has so far proved prohibitive. In a follow-up interview, a researcher gave a counterpoint to the suggestion above, noting that simple encryption, that is, password protecting a laptop, was not effective and gave researchers 'a false sense of security'.

Whilst there may be a requirement to preserve confidential data appropriately, it clearly cannot be openly shared. Open Access metadata-only records with related data securely stored in a publicly-inaccessible dark archive is a solution that will be explored by Open Exeter.

3.5 Research Methodologies

Chart 3.5.1 illustrates the types of methodologies most commonly used by researchers at Exeter. The number of researchers using observational methodologies was highest. Of particular relevance is the large number of researchers who use surveys, interviews, focus groups, and so on, all of which methodologies may produce data that has specific ethical, legal and storage requirements. It is clear that such researchers would benefit from core training in planning data security and storage.

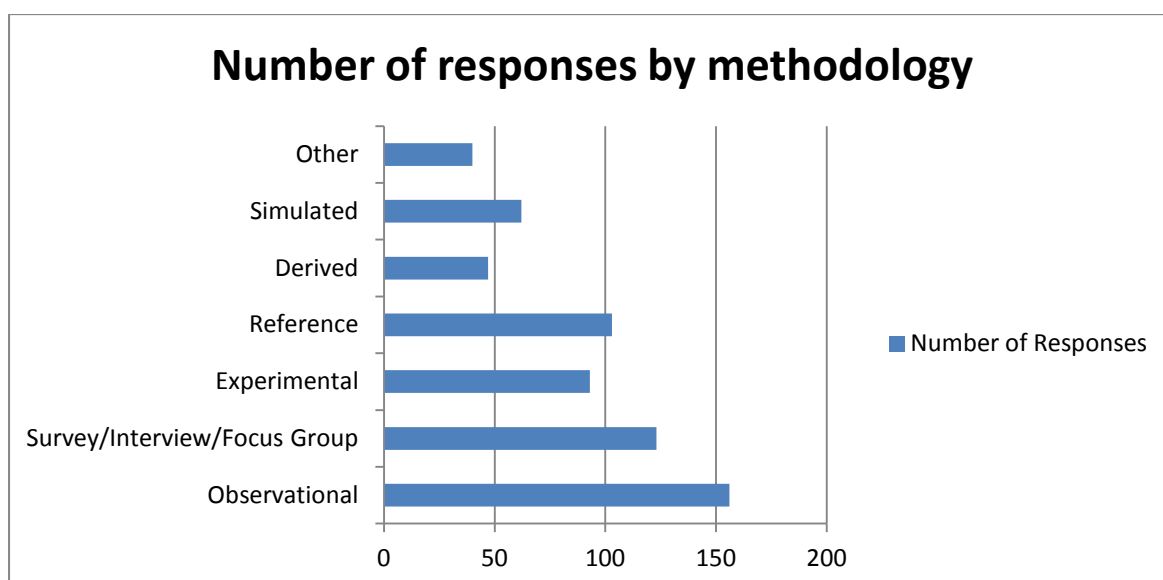


Chart 3.5.1: Number of responses by research methodology

When broken down to College level, results in this section provide a clearer picture. As can be seen from **Charts 3.5.2 and 3.5.3**, it is clear that broadly speaking different Colleges employ different methodologies. This finding reinforces

awareness that Open Exeter training sessions must be made relevant to each College or discipline. Subject-specific training has been mentioned many times in interviews - even College-level PGR training on research skills is sometimes perceived to be relevant to only a few attendees, by focusing, for example, on generic qualitative or quantitative analysis skills and software.¹²

What is particularly striking is the percentage of researchers in both CSSIS and The Business School who collect data via surveys, interviews or focus groups. As noted above, this result implies that training on handling and storing confidential or personal information will be a priority for these Colleges.

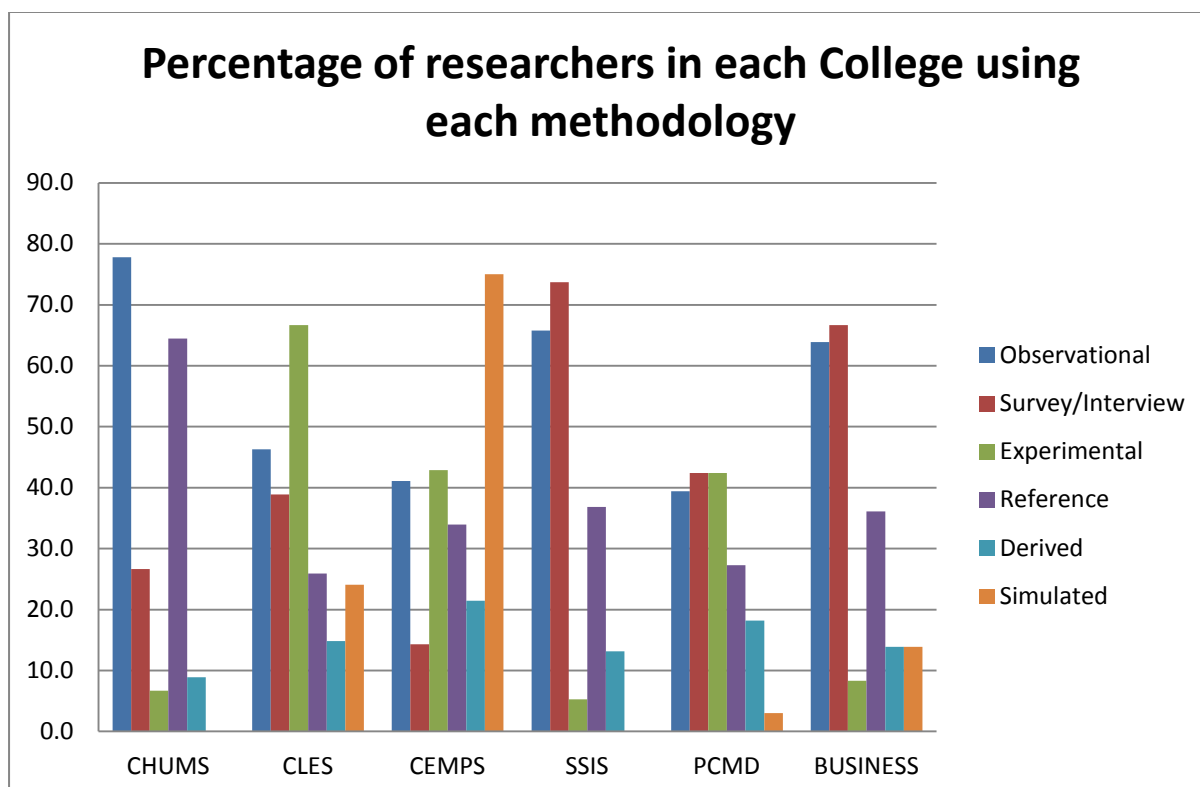


Chart 3.5.2: Percentage of researchers, by College, using each methodology

¹² For more on training needs see Section 8 of this report.

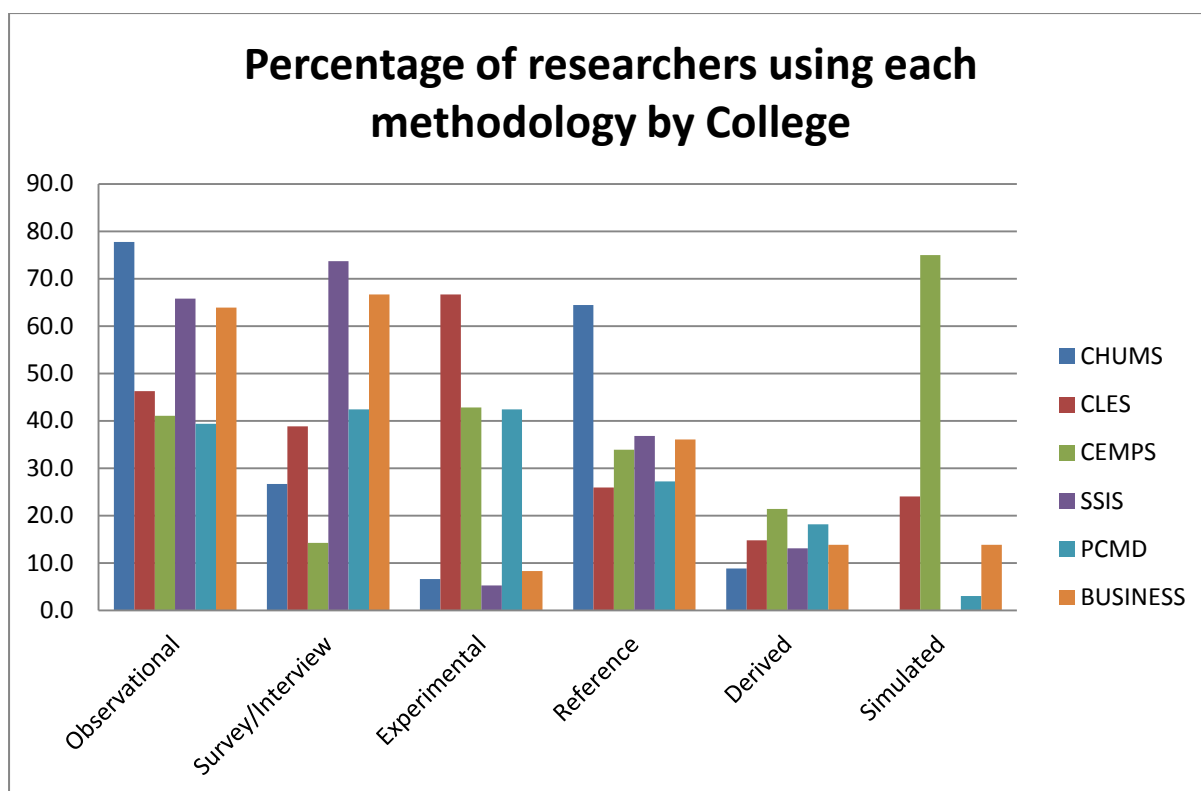


Chart 3.5.3: Percentage of researchers in the Colleges using each methodology

3.6 Data Types

Charts 3.6.1-3 reveal percentages of respondents using each data type listed on the survey form. Documents and spreadsheets are the predominant file types used by researchers with the exception of CHUMS where use of spreadsheets is much lower (by 19% points). In reference to use of databases, all College results, with the exception of the Medical School (PCMD) at approximately 50% of respondents, are around 20%.

The percentage of researchers across the Colleges whose data incorporates images is noteworthy. The figure for CHUMS researchers is the highest and may be due to archive-based research generating increasing numbers of digital images, a supposition supported by evidence from interviews. For example, one researcher is currently storing 20,000 archival images. Proper storage and use of image data requires specific knowledge that will need to be built in to training where appropriate.

It is revealing that the chart for those researchers who collect audio data is almost identical in shape (although at a lower level) to that of researchers who use surveys and interviews in their research. Evidence from Open Exeter interviews suggests that most of the audio files created by researchers are generated during interviews, which may explain the similarity of the charts.

When looking at the combined data type chart (**Chart 3.9**), the use of 'Raw Data' appears to be the fourth most used data type. However, when broken down by College it is clear that 'Raw Data' is actually used only by CLES, CEMPS and the Medical School. As the term 'Raw Data' may have a number of different meanings,

the survey help hinted that 'Raw Data' might, as an example, be defined as 'Machine generated'.

Following feedback from the survey pilot, an additional data type 'Unique programme created specifically for the project' was added. It was assumed that this data type would be used predominantly by CEMPS and, to a lesser extent, CLES researchers. However, a breakdown of the results shows that this is not the case. Although nearly 50% of CEMPS researchers state that they do use a unique programme, so do nearly 20% of Business researchers. The figure is only just below that of CLES researchers. It will be interesting to discover through interview research whether new programmes are being written for/by Business researchers or whether there was a misunderstanding of the question being asked.

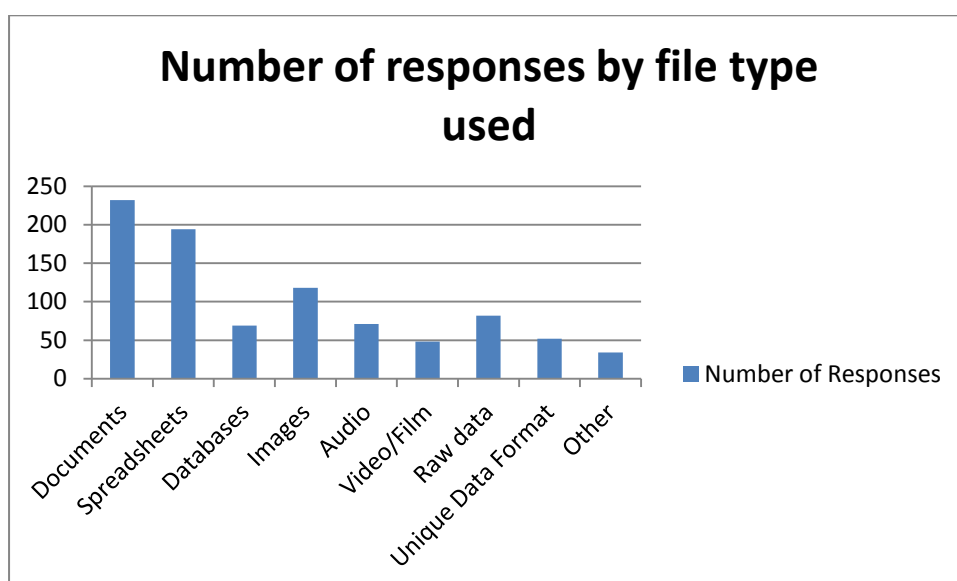


Chart 3.6.1: Number of responses to the survey by file type used

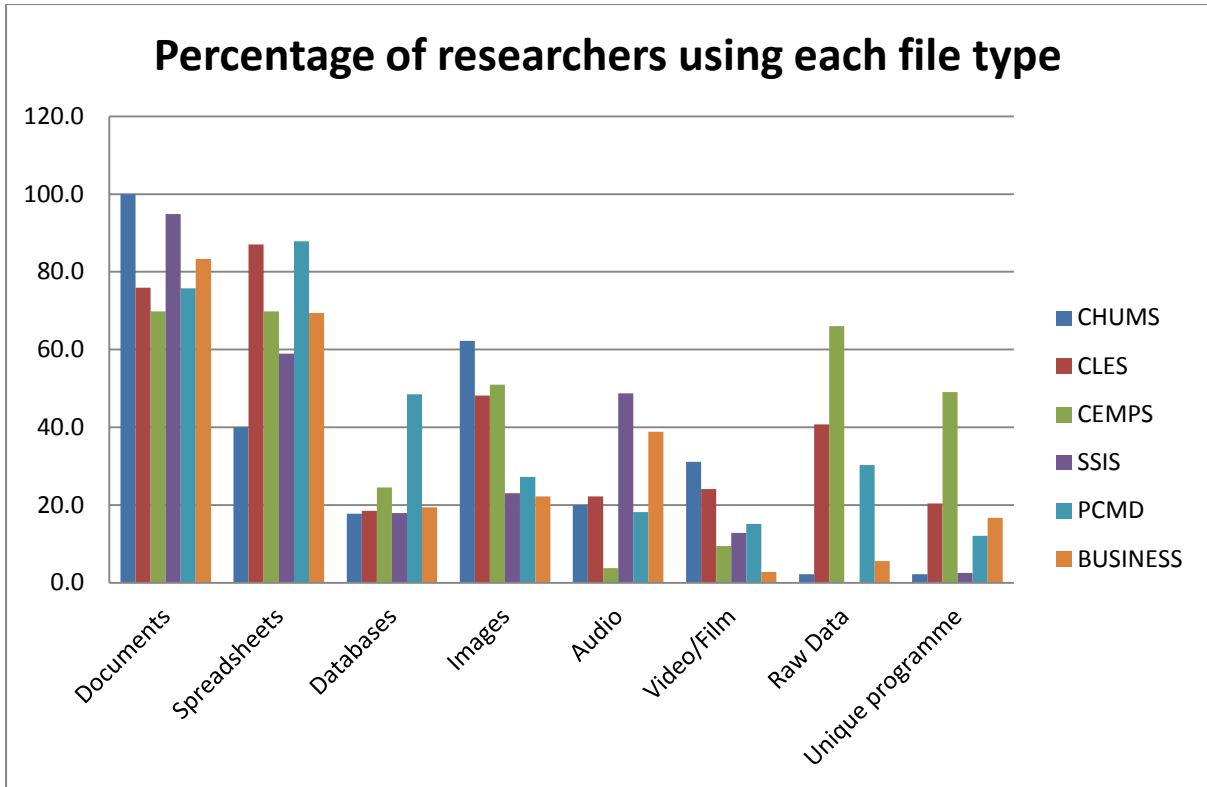


Chart 3.6.2: Percentage of researchers using each file type

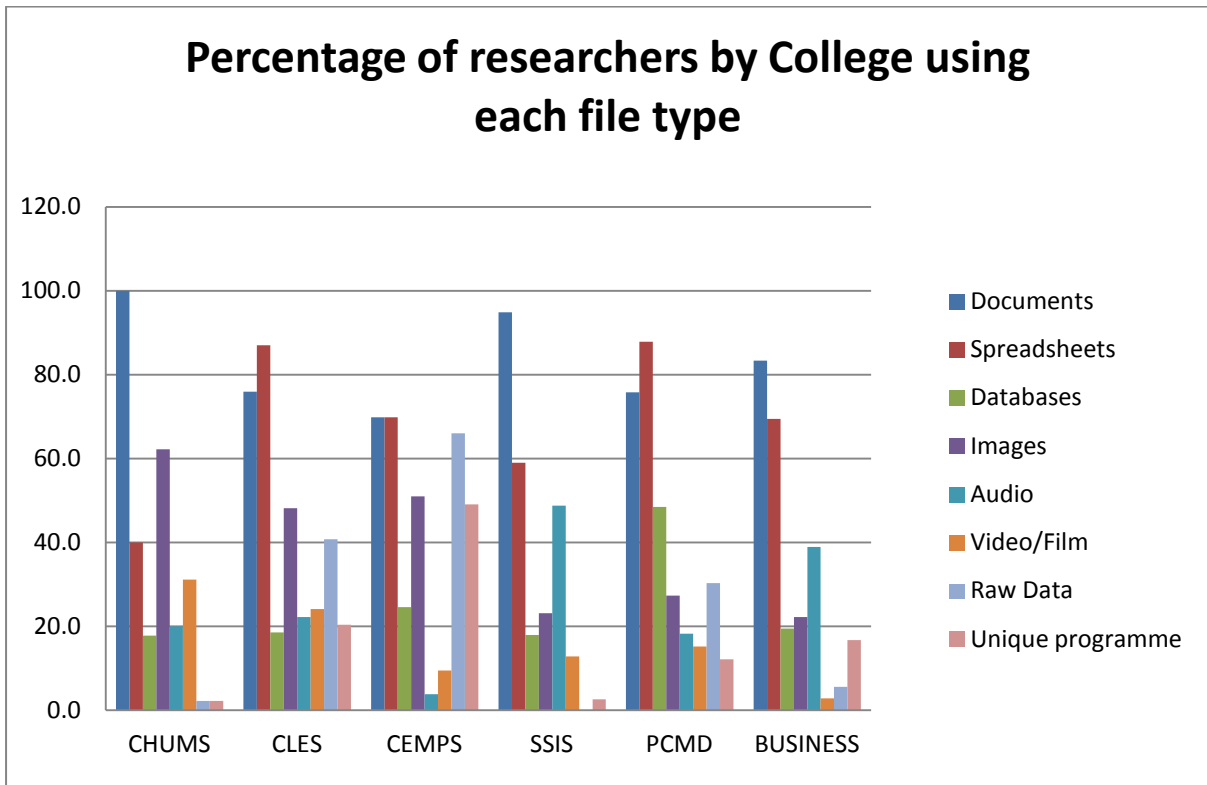


Chart 3.6.3: Percentage of researchers, by College, using each file type

The range of data types used across the University suggests that while EDA must be file agnostic, at the same time researchers must be encouraged, through training

and guidance, to favour open, standard formats that will facilitate sustainable long-term preservation and data curation.

3.7 Data Storage

3.7.1 Live Data Storage

Approximately half of survey respondents used their office computer's hard drive to store at least part of their live data (**Chart 3.7.1.1**). Half of the respondents also used the University Network Drive. However, these figures do not give a full picture of current storage practices. The University provides 20GBs of network storage to academics and PGR students which is backed up daily. In addition, the vast majority of researchers are provided with an office PC. A number of researchers do not, however, appear to use these University facilities. It should be noted at this point that there is a substantial number of Mac users throughout the University, in some cases entire departments, such as Drama. It is current University policy not to support Mac usage through the central IT support service. This decision clearly has an impact on Mac users and the options that are available (or in many cases, not available) to them for storage, sharing and transferring data. In hindsight, it would have been useful to try to obtain information about the number of Mac users and this is an area that may be investigated in future.

When asked "Where is the research data you are currently working on primarily held?" 24 of the 46 CHUMS respondents did not choose either their office computer's hard drive or University Network Drive as an option. This figure equates to over 50% of the respondents in CHUMS not using University-provided equipment to store live research data; follow-up interviews have supported this finding. One researcher hypothesised that as Humanities researchers have not traditionally conducted server-based research they are not accustomed to it. The same researcher also stated a preference for conducting research at home where there were fewer distractions. This lecturer separated teaching (on campus) and research (off campus). In contrast, the figures in the other colleges are much lower and can be seen in **Table 3.7.1.1**. Unlike scientific research, which is frequently lab based, Humanities research can often be carried out in any environment.

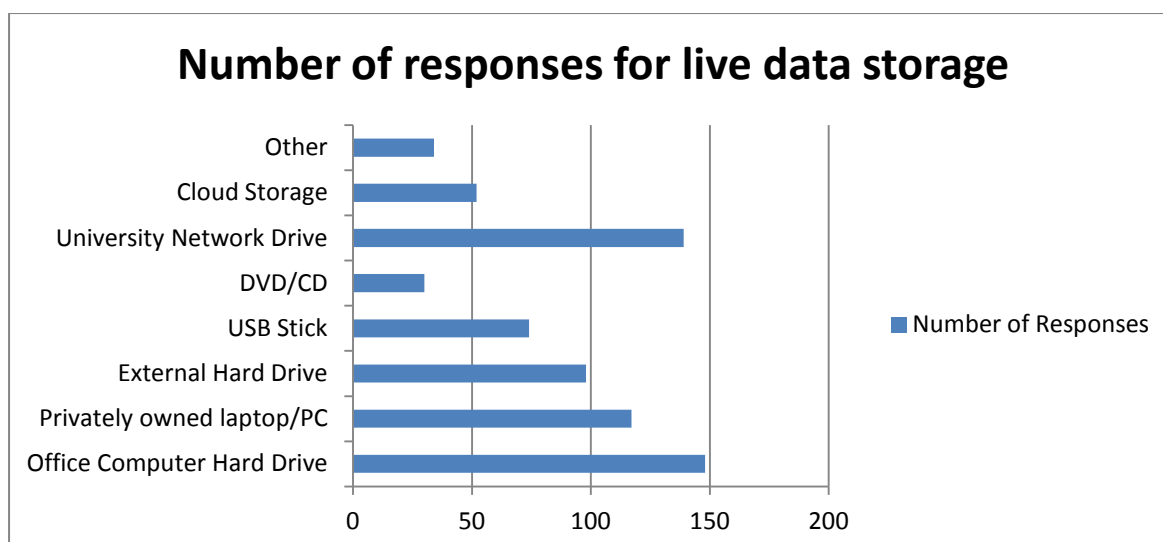


Chart 3.7.1.1: Number of responses to the survey showing where researchers “live” data is stored

College	Number of respondents not using either their office computer or the network drive	Percentage (%)
CHUMS	24/46	52
CEMPS	3/55	5
CLES	9/59	15
SSIS	14/41	34
Medical School (formerly PCMD)	7/46	15
The Business School	4/37	11

Table 7.7.1.1: Table showing the number and percentage of respondents who do not use either the University Network or their office PC to store their live research data

Precise reasons for lack of use of University facilities need to be investigated further. Interviews and group discussions suggest a combination of reasons, including: lack of awareness, lack of compatibility (if Mac users, for instance), not adequate for needs. It is clear, however, that awareness and appropriate use of University facilities needs to be introduced to all new staff and students at the earliest opportunity.

The percentage of researchers across each College using cloud storage for live data is illuminating (see **Chart 3.7.1.2**). In the Medical School, only a single researcher used cloud storage for their live data. Medical School researchers tend to be working with sensitive and confidential data which is not compatible with the current generation of cloud storage devices and companies. This implies that any future storage solution provided by the University should take the restrictions applying to Data Protection Act-covered data into consideration.

In interviews, one researcher who used health data stated that they did not use cloud storage because of the perceived security risks associated with storing confidential data in externally-hosted cloud services. When asked if they would consider cloud storage if the data was hosted at Exeter, they stated they would consider this option. However, they also stated that they would have to “trust” the solution. In addition,

Medical School researchers have noted that large datasets are not compatible with cloud storage solutions because of the time taken to upload files. However, researchers in the two Science Colleges make regular use of cloud storage so this may not be the complete answer.

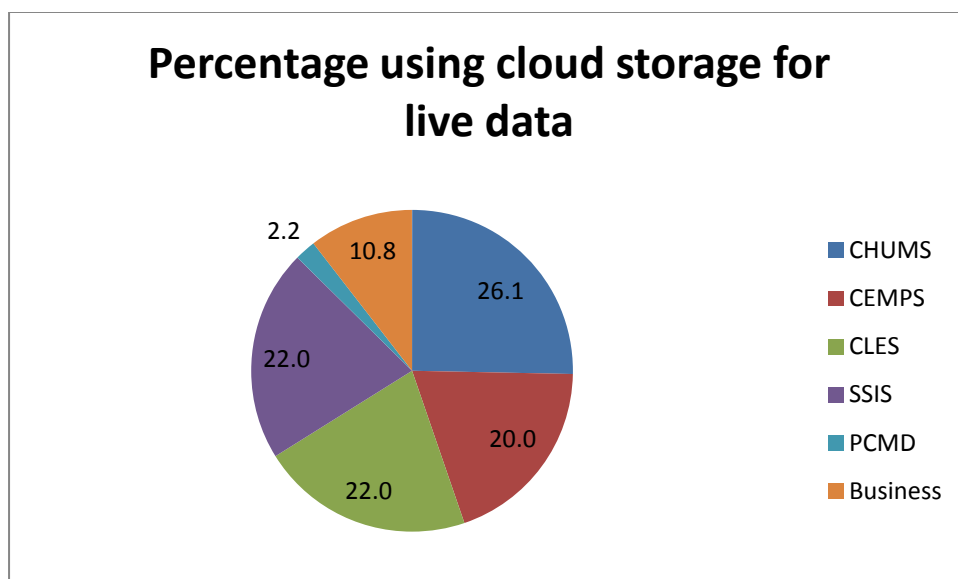


Chart 3.7.1.2: Percentage of researchers using cloud storage for live research data

Survey analysis provides some evidence that it is earlier career researchers who use cloud storage systems most frequently. Thus, of the nine users in SSIS, all bar one (a senior lecturer) described themselves as either a post-graduate or a post-doctoral researcher. In CHUMS, none of the professors used cloud storage and in CEMPS, only PGRs and post-doctoral researchers used it. The situation is different in CLES, where at least one respondent at each level used cloud storage and two of the four professors did.

As storing and sharing data quickly and easily via cheap solutions such as Drop Box becomes more common, it is clear that the University should, at least, release some basic guidelines on proper and appropriate use of cloud-based storage.

3.7.2 Size of Data

Size of data is an important element to be considered in the planning and provision of storage systems: the larger the data, the higher the associated costs. This proved to be a difficult question to phrase correctly in order to get the right sort of information from respondents. The question was an open one but some guidance was provided to encourage respondents to answer in terms of MBs, GBs and TBs.

Answers (where known/given) ranged from 0 (data collection not yet commenced) to 50TB. A large number of respondents did not know how much data they had, or answered generally, for example, "Recently bought our own server for large datasets"; "That is very hard to quantify. We have very large files containing the databases of electronic searches used to identify the relevant literature. The files containing the pooled and synthesised data will be small"; "Impossible to estimate. I store hundreds of coloured images which are always large files in terms of GB"; "A

lot. Too much to store on the university network drive, and enough to store on a 500GB hard drive”.

Researchers in a number of departments and Colleges have very close links with College-specific IT support (Computing Development Officers) who are able to provide guidance on data management issues, such as supplementary storage and back up. Interview findings suggest that access to a local named contact or contacts able to provide immediate, subject-specific support works well and is preferred by researchers. It was noted that this contact could be via email or over the ‘phone, this latter point being made by researchers at the Tremough campus in Cornwall whose support is based at Exeter.

The number of researchers who did not know the amount of data currently held (or related answers) was quite large: 78 out of 284, or 27%. However, as can be seen from **Table 3.7.2.1**, the number of unknowns varied widely between the different Colleges.

College	Number of unknowns	Percentage of respondents (%)
CEMPS	3/55	5
CHUMS	25/46	54
CLES	6/59	10
SSIS	14/41	34
PCMD	18/46	39
Business	12/37	32

Table 3.7.2.1: Number and percentage of researchers who do not know how much electronic research data they have

Some tentative conclusions can be drawn from this table:

1. Researchers in CLES and CEMPS tend to know how much research data they hold. This could be because they are aware of how to find this information, or that because they regularly generate large amounts of data, they are more aware of the need to monitor storage quotas.
2. Over half of CHUMS respondents do not know how much research data they hold. This could be because they do not know how to check for the size of their files, or because data files are typically of smaller size, not requiring additional storage.

These figures for CHUMS do conflict somewhat with what is known from interviews and group discussions: that CHUMS researchers, particularly those engaged in Digital Humanities projects, possess high levels of technical competencies, and regularly produce GBs of data. It could be that our survey did not reach or appeal to this group of researchers, and it will be important to examine this finding more closely.

The mean average of each College using the known responses is useful (**Table 3.7.2.2**).

College	Mean average file size (in GB)
CEMPS	1699
CHUMS	78

CLES	2632
SSIS	51
Medical School (PCMD)	1225
Business	52

Table 3.7.2.2: Table showing the mean average file size in each College

The considerably larger file sizes held in the Science Colleges are evident. The largest amount of research data from a single researcher in CHUMS was 500GB. That in CEMPS was 30TB and in CLES 50TB. Eleven CEMPS researchers had more research data singularly than the combined total of the 21 known CHUMS researchers.

However, these disparities should not disguise an important discovery. The University provides 20GB of backed up storage space for researchers. Even in SSIS (which has the lowest mean average of data size), seven of the 27 researchers where numbers are known held 20GB or more of data. Thus, even in the College with the lowest mean, over a quarter of researchers surveyed had more data than the space the University provides. In Business the figure was one quarter (5 out of 20) and in Humanities nearly half (10 out of 21).

It may be worth highlighting at this point a recent snapshot survey conducted by Simon Hodson, JISC Managing Research Data Programme Manager, comparing data storage systems in Russell Group and 1994 universities.¹³ Although both generated large amounts of research data (approximately two petabytes held in one Russell Group university), it was the Russell Group institutions that provided central data storage systems for researchers.

The results of this survey will be shared openly with Exeter IT and Colleges to help inform future strategies for live research data storage.

3.8 Data Back Up

Backing up is a vital component of effective research data management and it does appear that a majority of Exeter researchers are aware of this fact.

A number of interviewees have stated that students embarking on doctoral-level study are expected to know that they need to back up their work and data, if only because it is likely that they will have suffered data loss during the course of their studies. In the words of one researcher, "Who hasn't suffered data loss?" Whilst probably true in many cases, such assumptions could perhaps lead to a lack of back up knowledge and skills in new PGRs.

Only ten of the respondents stated that their data was not backed up. However, of these ten, a number also answered the question, 'Where is data backed up?' It is possible that some respondents gave different answers in reference to different elements of their research. For example, a large number of respondents stated that

¹³ <http://www1.uwe.ac.uk/library/usingthelibrary/servicesforresearchers/datamanagement/managingresearchdata/conference/programme.aspx>

they manage the back up of their data in addition to stating that somebody else manages back up on their behalf. This apparent contradiction can be explained if the different elements of their data are stored in different places, which the conflicting answers in the data storage section show to be the case.

It may also be the case that the concept and mechanisms of back up are not properly understood by all respondents. In addition to those who replied that their data was not backed up, a number of respondents stated that they did not know where or how their data was backed up. One Associate Professor who answered in this manner added the comment, “I assume everything on the shared drive/network is backed up, in which case it’s [backed up] however often that is”.

Researchers storing non-electronic research data (such as paper records, log books, magnetic tape) typically do not have this data backed up in any form. This finding emphasises the importance of a holistic approach to developing training and advice for research data management. For example, one single Associate Professor has, “...several files from over 20 years. Also four drawers of filing cabinets, 2 plastic storage boxes, and several box files.” In other areas there are “personal research notes”, “audio-cassettes”, “Fieldwork notes”, and so on.

Project interviews, typically conducted in a researcher’s office have led to the discovery of cupboards, cabinets, drawers and boxes containing a range of data of varying types, size, age, and format, some verging on obsolescence, and usually unique copies.

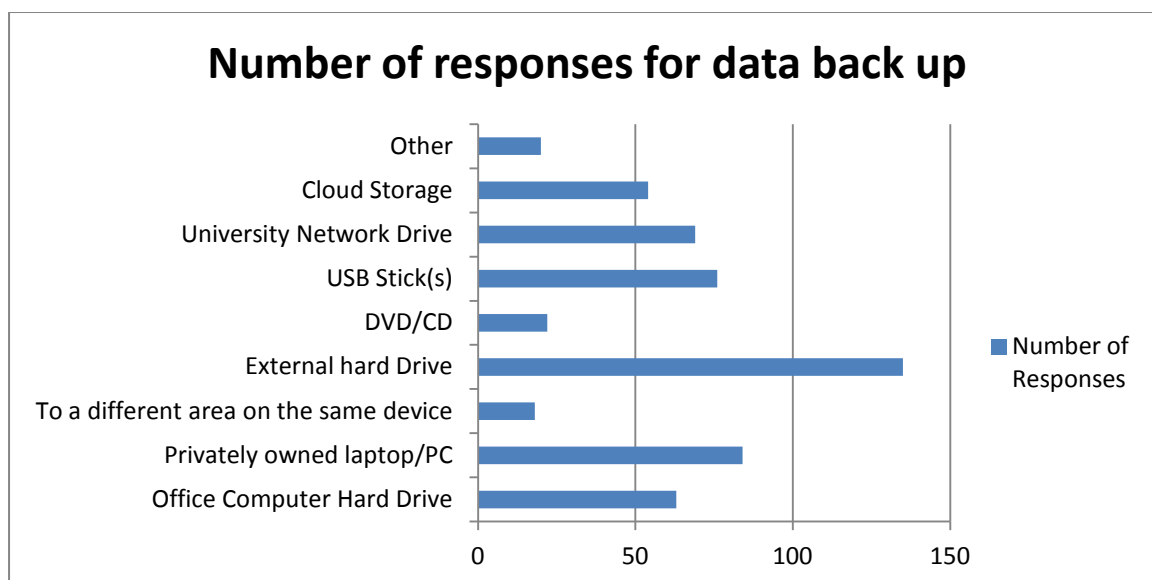


Chart 3.8.1: Number of responses for data back up

When broken down into College responses, it is clear that researchers in the different Colleges use the tools available to them for back up in different ways (**Chart 3.8.2**). Thus, the percentage of CHUMS researchers who back up their data to their office computer is substantially less than any other College. This finding supports evidence in the previous section that CHUMS researchers use their office computers less frequently than researchers in other Colleges.

In addition, the Medical School and SSIS researchers do not use cloud storage as a back up option as much as researchers in other Colleges.

Use of external hard drives for back up is common. This finding is supported by results emerging from the Open Exeter Follow the Data initiative involving PGR students from the different Colleges. All of these students use external hard drives to back up their own work. One Open Exeter PGR noted that at a previous University external hard drives had been loaned out to PhD students to aid them in their work. Certain Colleges could consider this where appropriate or financially viable. However, external drives are at risk of failure, cannot be relied on completely, and are effective only if back up is regularly and methodically carried out. University-wide guidelines on appropriate use of external devices (including what types of data should and should not be stored on portable devices) would clearly be useful.

PGRs can be viewed as a specific target group as they are, perhaps, at greater risk of losing data. As PGR work is frequently embedded in and essential to project and research group work, the management and preservation of their data should be given serious attention.

For note, the percentage of researchers using DVDs and CDs to back up is very small across all Colleges.

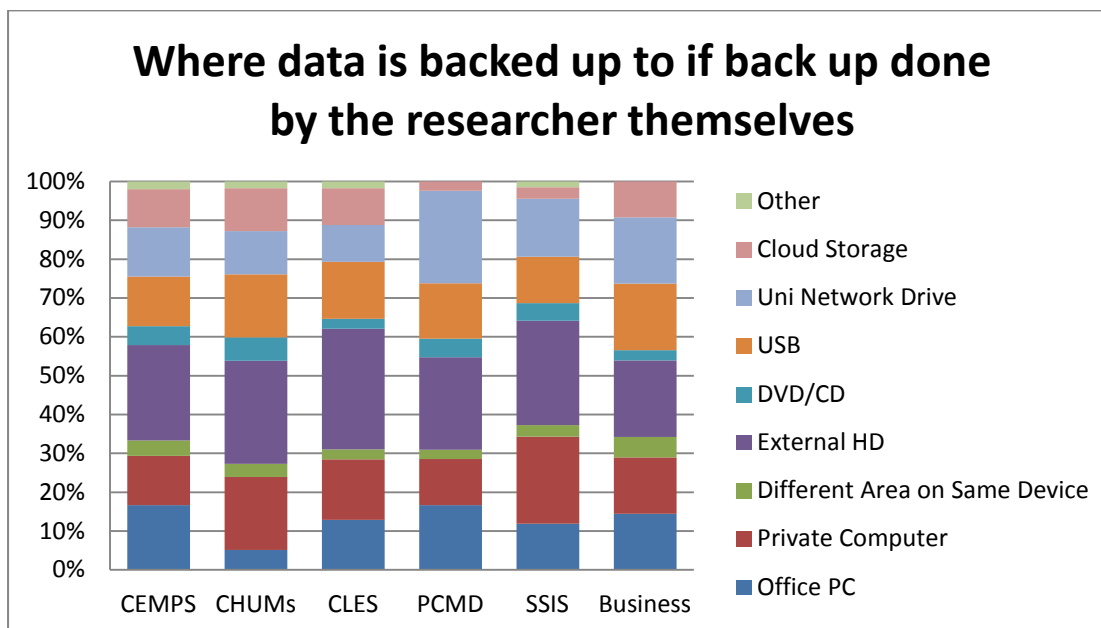


Chart 3.8.2: Graph showing where researchers back up their data, if they back up themselves

The Medical School has the greatest proportion of researchers who do not back up their own data. 16 Medical School respondents (out of 46) stated that somebody else backs up their data (IT staff or automatic back up, for instance).

Meanwhile, CHUMS and SSIS researchers are most likely to back up data to a private computer or laptop.

Another interesting finding is that very few researchers back up only to one place. **Table 3.8.1** gives the mean average for those researchers who manage their own back up and who know exactly where data is backed up.

College	Average number of places backed up to by researchers
CEMPS	2.1
CHUMS	2.7
CLES	2.4
Medical School (PCMD)	1.7
SSIS	1.9
Business	2.2

Table 3.8.1: Table showing the mean average of places where researchers back up to

Researchers at Exeter are clearly aware of the importance of back up and, generally, work is backed up to more than one location. Questions remain, however, around an over reliance on vulnerable external devices, and the amount, quality and consistency of advice given to PGRs and (information gathered from interviews) new research staff.

The survey subsequently addressed the topic of data loss. Analysis of this section is being conducted separately and a report, possibly for internal use only, will be produced based on that analysis. Although a large proportion of researchers had suffered some data loss, the vast majority had lost only a comparatively small amount, such as a single day's work. Interviews have suggested that many researchers become more aware of the issues surrounding data loss when they have been personally affected.

Establishing a culture that promotes good practice in RDM is clearly an important factor in preventing any data loss. The consequences of large scale data loss to an externally funded project could be severe in terms of loss of reputation, inability to achieve compliance with funder policy, and potential loss of future income.

Whilst practice varies widely according to discipline, the notion of managing the materials that underlie research is not new to researchers. Using the right language and understanding the discipline will be key to increasing confidence and competency in the face of new funder requirements and the formal completion of data management plans.

3.9 Data Archiving

The phrasing of questions around the topic of data archiving proved difficult to finalise. Opinions on what constitutes the process and purpose of archiving differ across disciplines. It is known that the very word 'archive' is off putting to some researchers and that other researchers do not classify themselves as users of 'data'. Question wording was devised to be as straightforward and clear as possible, so that responses obtained would be of a kind as much as possible.

Responses received in this section suggest that there is indeed some confusion regarding archiving and backing up. This finding is perhaps best demonstrated by responses to question 23: “How do you currently archive the important elements of your research data once you have finished with it?” Results are shown in **Chart 3.9.1**.

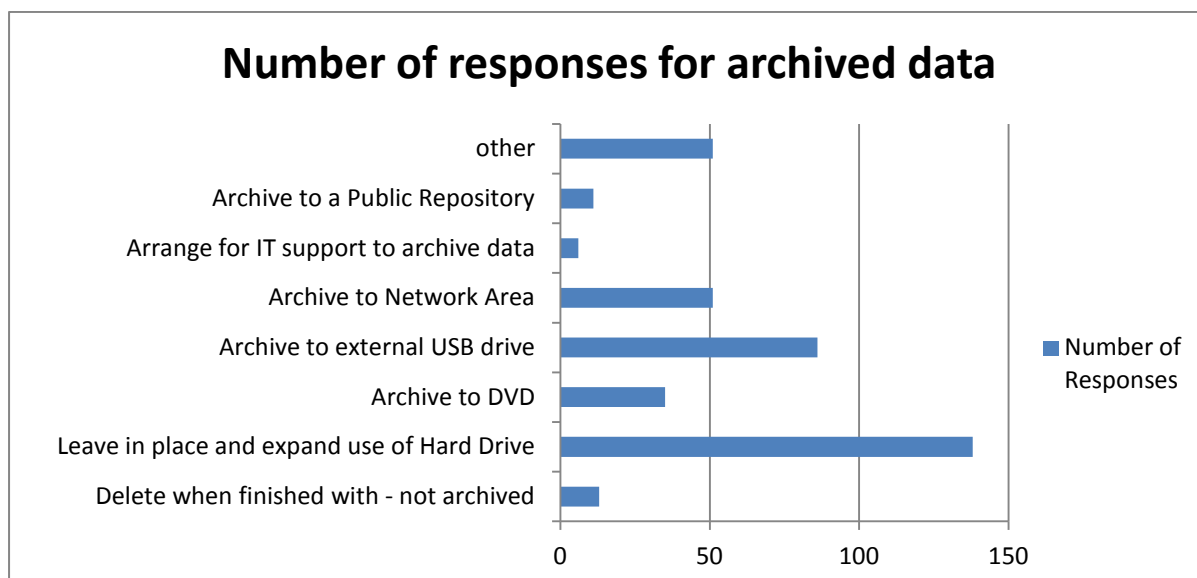


Chart 3.9.1: Number of responses as to where data is archived

By far the most popular answer was, “Leave it in place and expand use of hard disk”. In other words, data remained on the user’s computer/device rather than being transferred to an online archive or repository. Nearly 50% of respondents answered in this fashion. It is evident that archiving as understood by the Library and Data Curation communities is not a feature of most respondents’ research lifecycle and that increased understanding must be embedded in DMP training and support.

Perhaps a more helpful response can be seen in the answers to question 21: “How is it decided which of your research data is archived?” The categories in **Chart 3.9.2** have been collated from the 284 answers and there is only one response per respondent. As can be seen on **Chart 3.9.2**, the largest number of respondents archived (as they define the term) after they have finished a particular project, paper, model simulation, and so on. This habit may be expected, particularly where projects form discrete elements that are completed in a broadly linear style. Interviews suggest that data is frequently organised by research paper/publication. In addition, where active appraisal of data is carried out, data that has contributed to the completion of published research is preserved in preference to other data. These findings are exemplified by responses in this section.

However, even though the largest number of respondents answered in this fashion, if the answers to “Don’t know”, “N/A”, and “Don’t archive” are added together the total is 110 researchers (almost 40%).

This was a free text question allowing respondents to answer as they wished. Some of the answers follow:

- “I don’t know – I’ve just heard from my supervisor that it is” (PGR).

- “Usually I would archive data after I had finished publishing work on it. However, I have far less electronic and physical space/facilities here to do this compared with where I used to work, so I don’t know if this will be feasible. If there is nowhere safe to store data I would then have to destroy it” (Research Fellow).
- “It isn’t [sic] archived as such, just kept” (PGR).
- “I will hold electronically on external device indefinitely [sic]. Unsure how data in my university network is treated after I graduate” (PGR).
- “The data management act requires retention of interview data and this is specified in the informed consent form and ethics application (university). I have not yet archived data at this university or with this funder. Archiving should be decided as part of the research design process, but this is a fairly new way of thinking and acting, and not followed rigorously. I am sure we declined to deposit the data from this project with the ECSR because the interviews are sensitive in nature” (Senior Lecturer).
- “All the ‘raw’ data is immediately archived upon receipt. Archiving of data from the analysis stage is down to the specific researcher performing the analysis. We typically store the exact syntax used to generate the results, and also the results themselves” (Associate Research Fellow).
- “As far as I know, the data continues to be stored on the University shared drive” (Post-Doc).
- “We PDRAs have not been given any rules or even indications on this and our technicians have decided that two years (or maximum after publication) is sufficient archiving time” (Post-Doc).
- “I do not archive my data. All is under use or potential use” (Lecturer).
- “I am thinking about it now, which is why I agreed to do this survey” (Professor).
- “When the dust on my files is annoying my wife, I archive them in a different cupboard” (Occasional Teacher).
- “I am not sure of the details of this – would need to ask my supervisors” (PGR).
- “Most of my data does not need to be archived, it is sufficiently small to stay on my laptop” (Senior Lecturer).

As can be seen from these comments, the concept of archiving varies between researchers. The PGR student’s comment, “It isn’t archived as such, just kept”, adds credence to the view expressed above that a great deal of completed research data is simply kept on a hard drive, rather than archived in the ‘proper’ sense. Raising awareness of the purpose and benefits of data archiving is inextricably linked with a shift towards greater openness and sharing. Targeted advocacy that helps researchers to understand why they should archive (increased citation, funder compliance, easier collaboration, etc.) will be essential to ensure that the archiving process becomes a natural feature of data management.

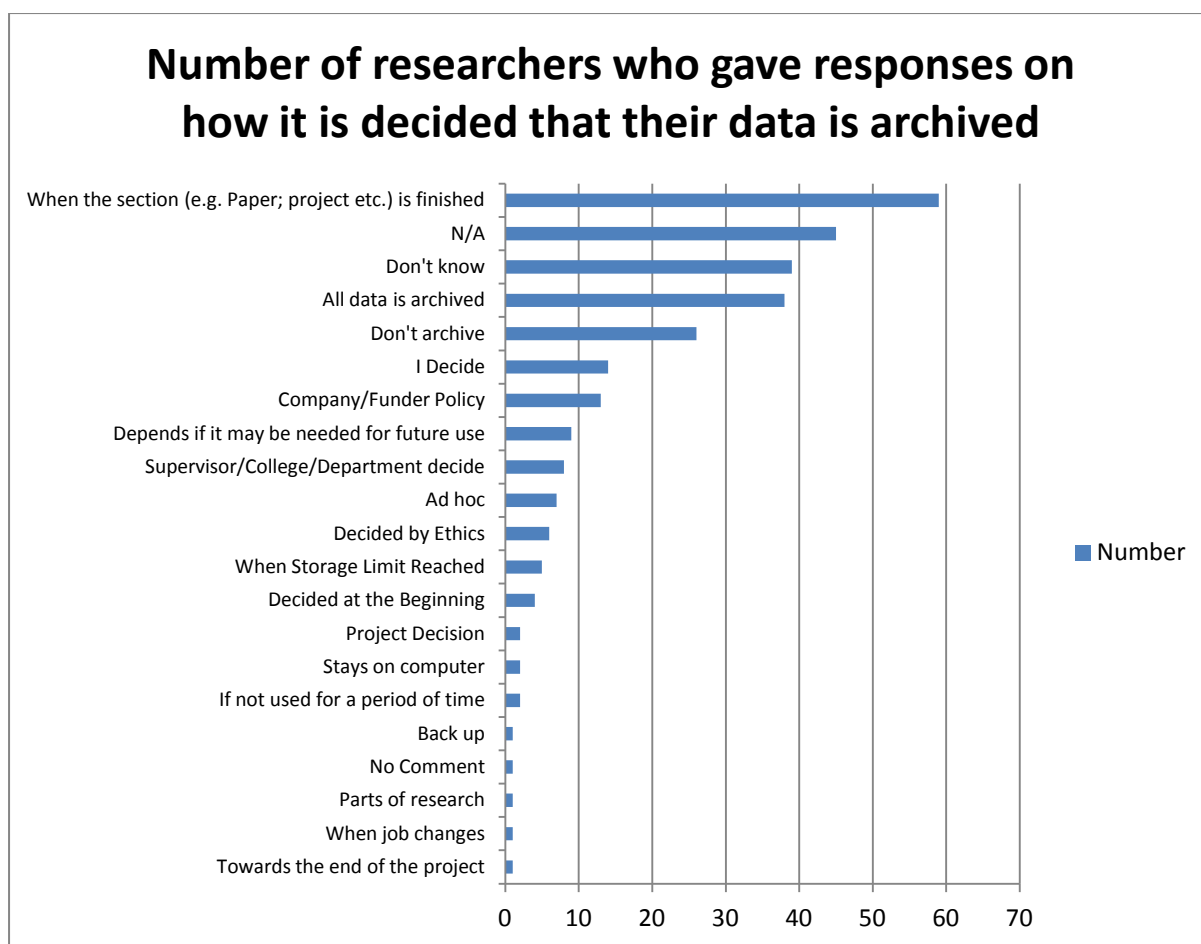


Chart 3.9.2: Graph showing responses as to how it is decided that researchers' data is archived

Some interviewees have mentioned that in addition to the actual data it would be helpful to archive software, virtual operating systems, and in some cases, hardware. However, institutional archiving of software or operating systems would involve licensing issues and permission would have to be sought in all cases. As in this instance, institutional solutions will not always be possible, emphasising again the importance of project-based data management plans addressing live and archive requirements associated with data specific to each project.

The variety of responses in this section shows that the language used when discussing RDM issues must be clear and understandable by all. Additionally, analysis of this section has shown that very few University researchers currently archive their data in the manner likely to be expected by all RCUK councils in the future (assuming they follow the EPSRC's lead).

Training in this area is an obvious priority and must be underpinned by the principles of the University's research data policy (currently in development). Clarification regarding when to archive and what to archive (criteria for selecting data for retention or disposal) will be important here.

3.10 Data Sharing

Chart 3.10.1 shows that almost two thirds of respondents (64.4%) share research data. However, when this number is split across academic Colleges it is clear that Science disciplines share research data more frequently.

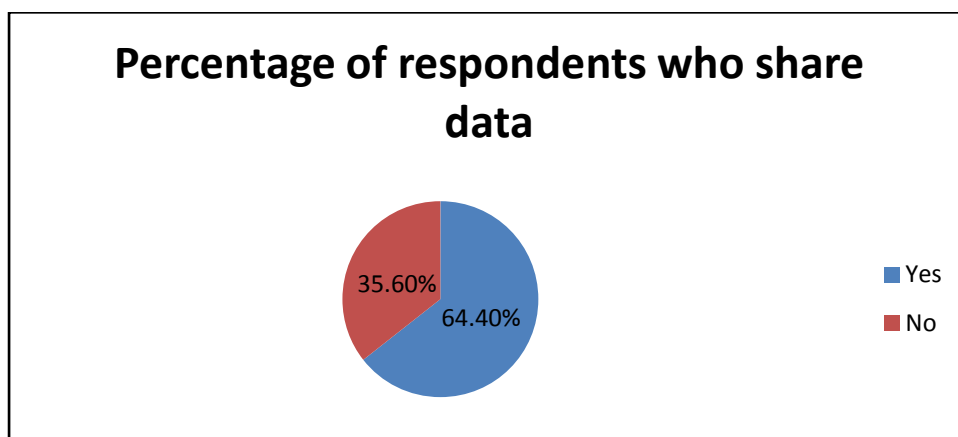


Chart 3.10.1: Percentage of respondents who share their data

Chart 3.10.2 depicts this finding quite vividly. 78% of CEMPS respondents shared data, 81% of CLES respondents and 80% of those in the Medical School (PCMD). In contrast, only 39% of HUMS respondents shared data, 46% in SSIS and 49% in The Business School.

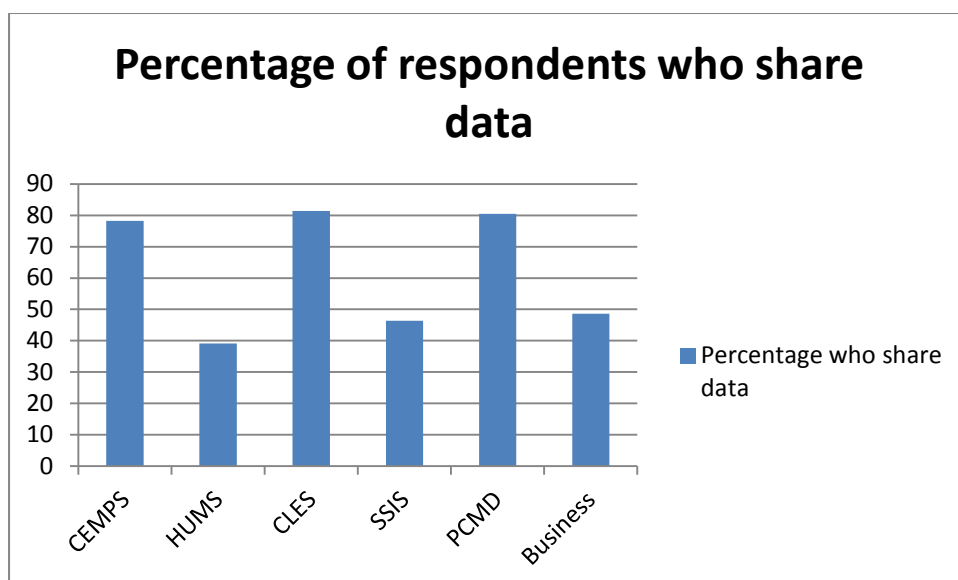


Chart 3.10.2: Percentage of respondents, by college, who share research data

Chart 3.10.3 shows that 52% of PGRs share research data. This percentage is substantially lower than all other positions with the exception of senior lecturers. There is a very understandable probable explanation: during interviews a number of researchers who supervise students have stated that they advise students to be cautious about sharing data too openly: not only may the student's doctoral work be at risk of plagiarism or unfair usage, but also the work of the associated research group could be compromised.

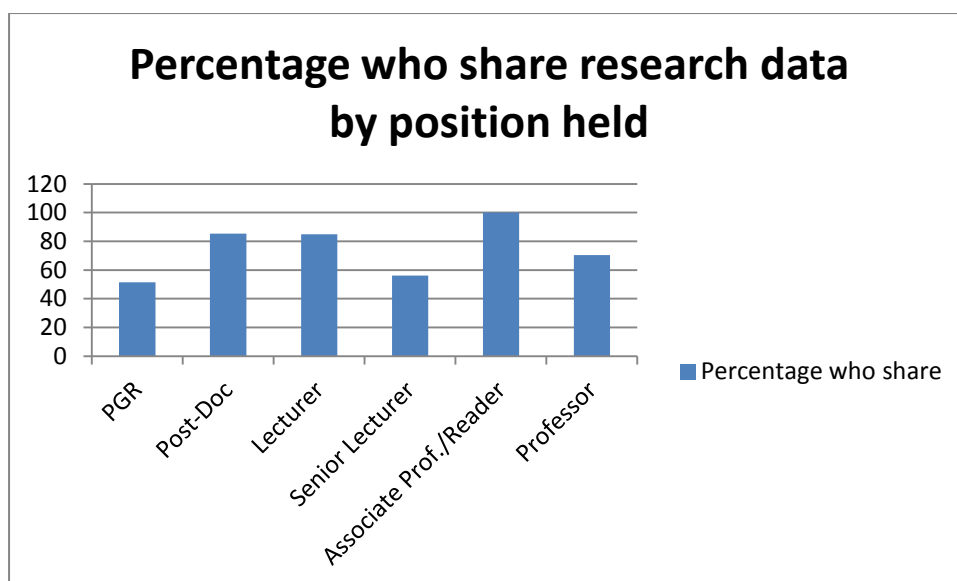


Chart 3.10.3: Percentage of respondents, by position, who share research data

How researchers share data is important. **Chart 3.10.4** shows that by far the most popular method of sharing research data was via email, with cloud storage and USB sticks distant second and third options. These methods have also featured in follow-up interviews. One researcher spoke of walking over to a colleague's office and picking up a USB stick in preference to the slow and frustrating process of downloading files over the network.

In addition, of the 183 researchers who do share data, only 34 use encryption for some of their data.

According to survey results, at Exeter, data is predominantly shared by email (noting the Exeter email servers only allow 50MB to be attached to each email). Use of the University network to transfer sometimes very large files, internally and externally, obviously has implications for network performance. Email cannot be relied on as a sharing mechanism as network and ISP quotas can block delivery. There are obvious security risks inherent in email transfer and this is clearly an unsuitable method of sharing any confidential data.

The University must provide clear guidelines and guidance on appropriate, ethical principles and methods of proper data sharing. Within an increasingly interdisciplinary and collaborative research culture, simple but secure means of sharing of data will become more and more important.

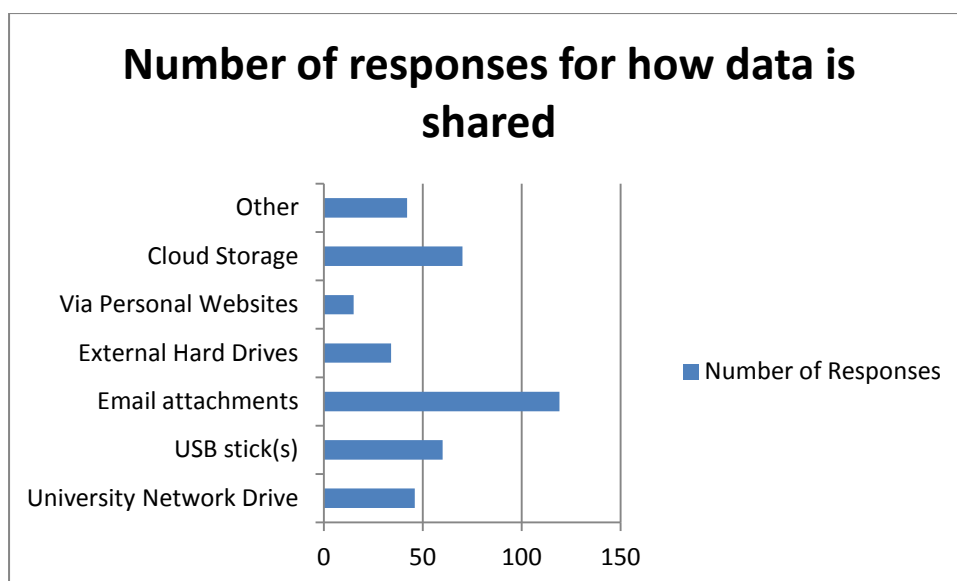


Chart 3.10.4: Graph showing how researchers share data

3.11 Training

An important output of the Open Exeter project will be a suite of training materials of various types and formats. For this reason, it was important to ask a range of questions focusing on training needs.

Only 31 respondents (11%) had received any data management training. Of these 31, eight had received their training from outside the University of Exeter (and one of the respondents was a PGR student working on the Open Exeter project).

Respondents were asked which training topics (selected from a list) they would prefer: How to Develop a Data Management Plan; Organising Research Material; File and Document Management; Legal and Ethical Issues; Bibliographic Software; and Institutional Repositories and Open Access. General results are shown in **Table 3.11.1**:

Training Area	Number
How to Develop a Data Management Plan	144
Organising Research Material	123
File and Document Management	112
Legal and Ethical Issues	115
Bibliographic Software	83
Institutional Repositories and Open Access	121

Table 3.11.1: Table showing what training respondents would like

These results show that the greatest number of researchers would like training in Data Management Plans, followed by Organising Research Material and Institutional Repositories and Open Access, with training in Bibliographic Software a distant last. Major funders now use evidence of planned data management as a criteria by which to assess funding proposals, therefore the high importance given to DMP training is not surprising. Accelerating awareness raising and training for data management plans and data management in general is a key recommendation of this report.

When findings are analysed by College the picture is more complex. **Chart 3.11.1** shows the percentage of respondents, broken down by College, who would like training in each area:

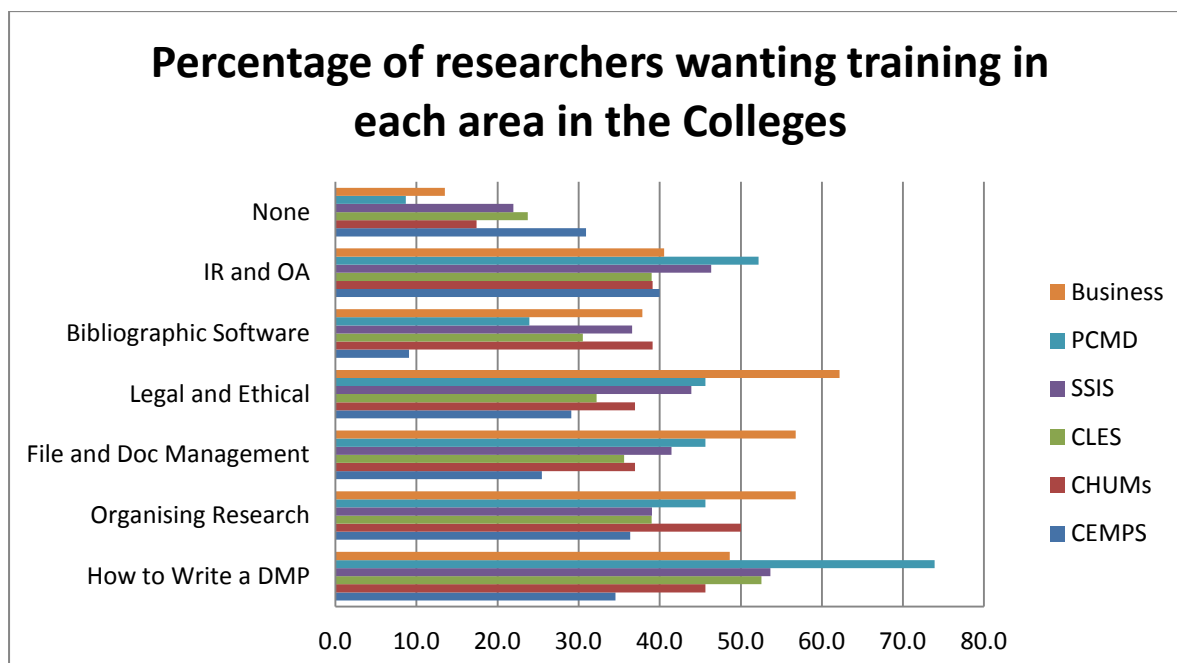


Chart 3.11.1: Percentage of researchers wanting training in each area

Three conclusions are immediately obvious from chart above: the percentage of researchers in the Medical School (PCMD) who would like training on data management plans is far higher than in any other College. Conversely, the percentage of researchers in CEMPS who want training in Bibliographic Software is far lower than in all the other Colleges. CEMPS also has the greatest number of researchers who do not wish to have any data management training, with the Medical School having the least.

The results per College make clearer the fact that fewer researchers in CEMPS would like any training (**Chart 3.11.2**). With the exception of Institutional Repositories and Open Access training the percentage of CEMPs researchers who would like training is lower across the board than in the other Colleges.

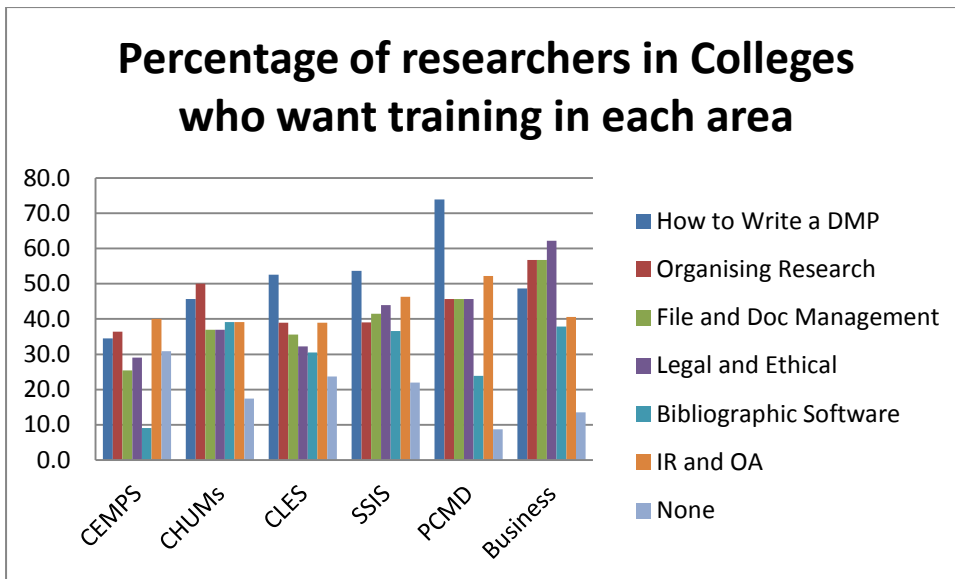


Chart 3.11.2: Graph showing what training respondents in each College would like

Open Exeter is developing training modules for the PGR Researcher Development Programme (RDP), delivered via workshops. Survey results show that training workshops and online training materials are the preferred learning methods of researchers at Exeter. **Chart 8.3** shows that this finding is consistent across all Colleges with the exception of The Business School where more researchers would prefer presentations and talks to online training materials. The findings of the DAF survey will help ensure training methods are tailored to disciplinary requirements.

Supplementing this finding is evidence from work with PGRs where it has been very apparent that an open workshop format comprising attendees from across disciplines encourages free, unconstrained sharing of experience, skills and knowledge from which all participants can benefit.

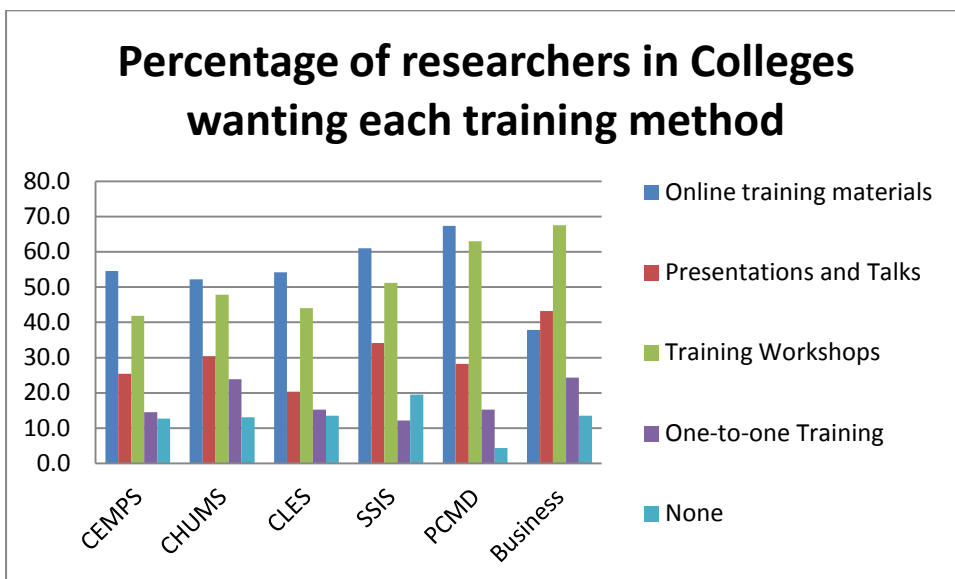


Chart 3.11.3: Graph showing, by College, what training method researchers prefer

The number of researchers who would like one-to-one training is lower than expected, actually the lowest of the four methodologies listed. This is shown clearly by **Chart 3.11.4**.

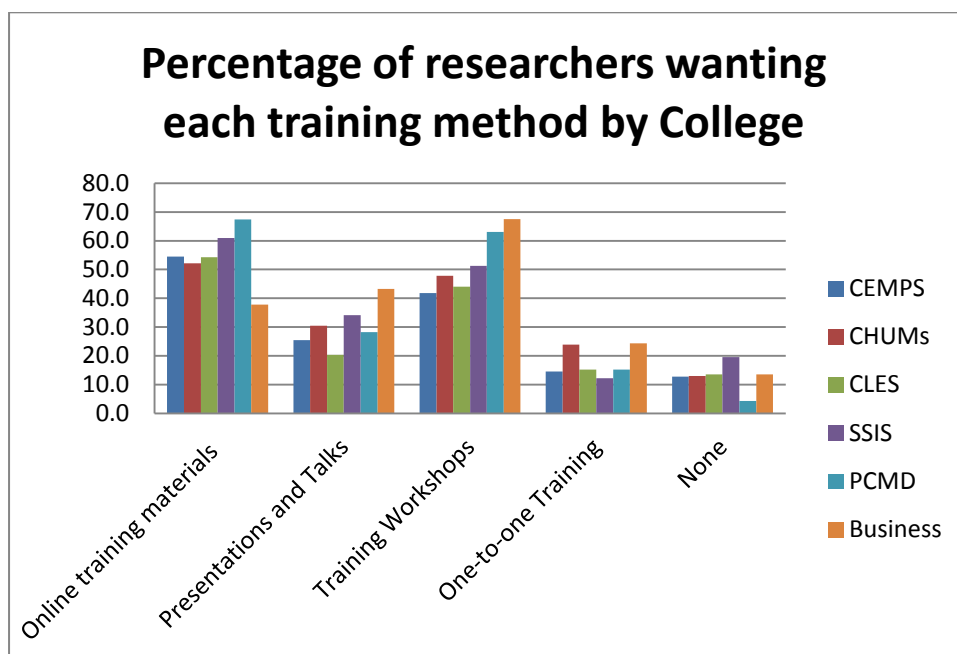


Chart 3.11.4: Percentage of researchers wanting each training method

As PGR students were the largest single group of respondents it was felt that an analysis of their training needs would be of use and interest. This would also help the Open Exeter project with its design of RDP training modules.

PGRs and early career researchers will play an essential role in the gradual embedding of an institutional culture of good RDM in the hope that they will carry good practice forward in their academic careers and pass on their knowledge to their students. The needs of PGRs require particular focus in the light of this report.

The percentage of PGRs who would like training in each area highlights some interesting points. Currently, there are RDP courses on the Data Protection Act and Endnote. The existence of these modules may explain the lower percentage of training required in these two areas (**Chart 3.11.5**).

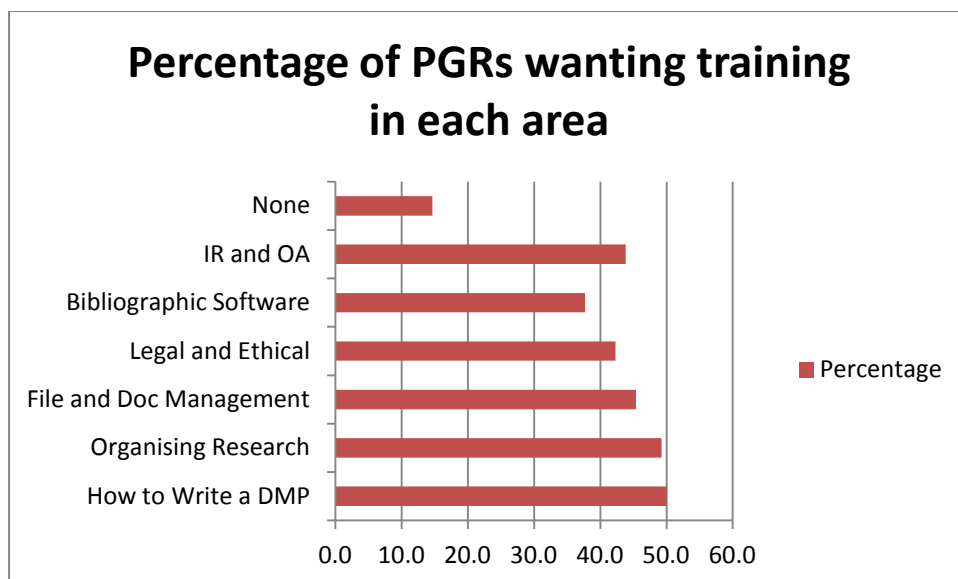


Chart 3.11.5: Percentage of PGR students wanting training in each area

As with all sections, closer analysis shows that there are differences between the academic Colleges. As can be seen in **Charts 3.11.6 and 3.11.7**, the differences between CEMPS and CHUMS are obvious.

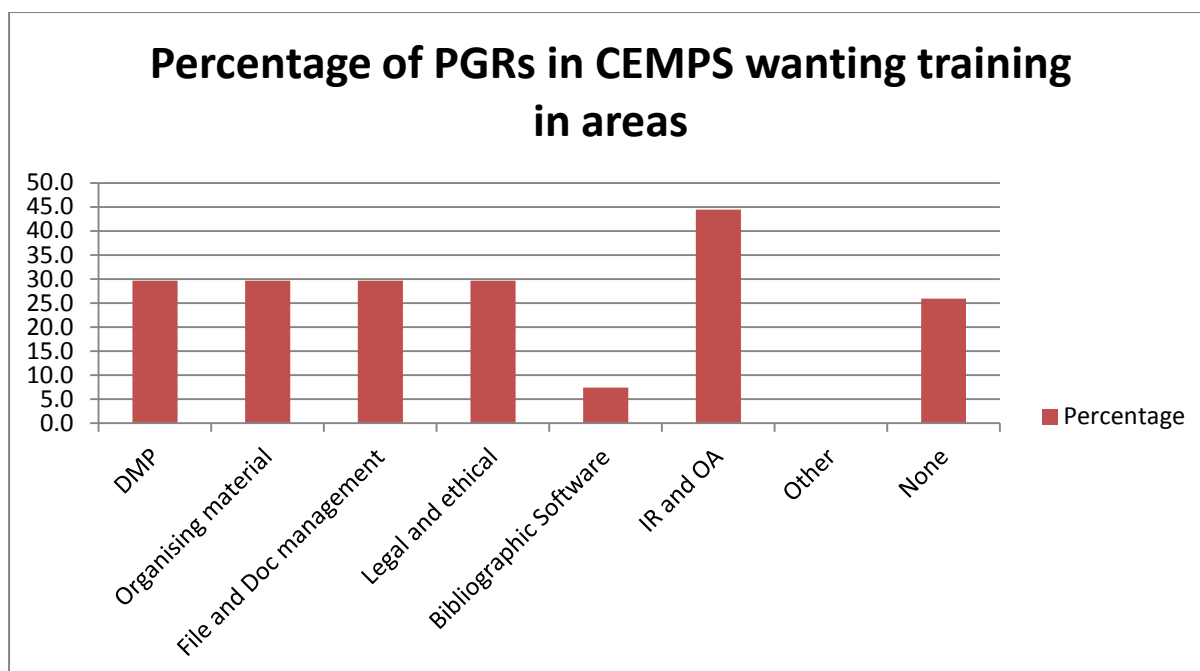


Chart 3.11.6: Percentage of PGR students in CEMPS wanting training in each area

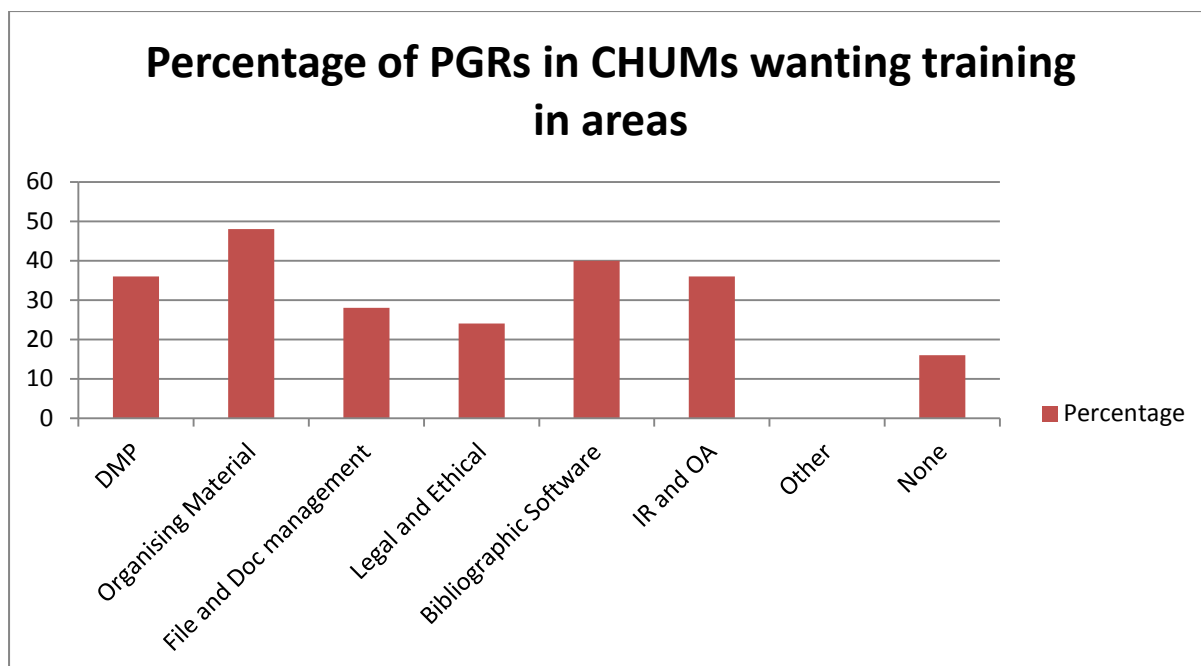


Chart 3.11.7: Percentage of PGR students in CHUMS wanting training in each area

How Exeter’s PGR students would prefer to be taught is interesting. Frequent reference is made to the current cohorts of students being “digital natives” and preferring all resources to be online. However, these results show that although a large percentage do like online training materials, the highest percentage prefer training workshops. Again, this finding may be related to what is available via current RDP courses. One-to-one training is by far the least preferred of the four methodologies (**Chart 8.8**).

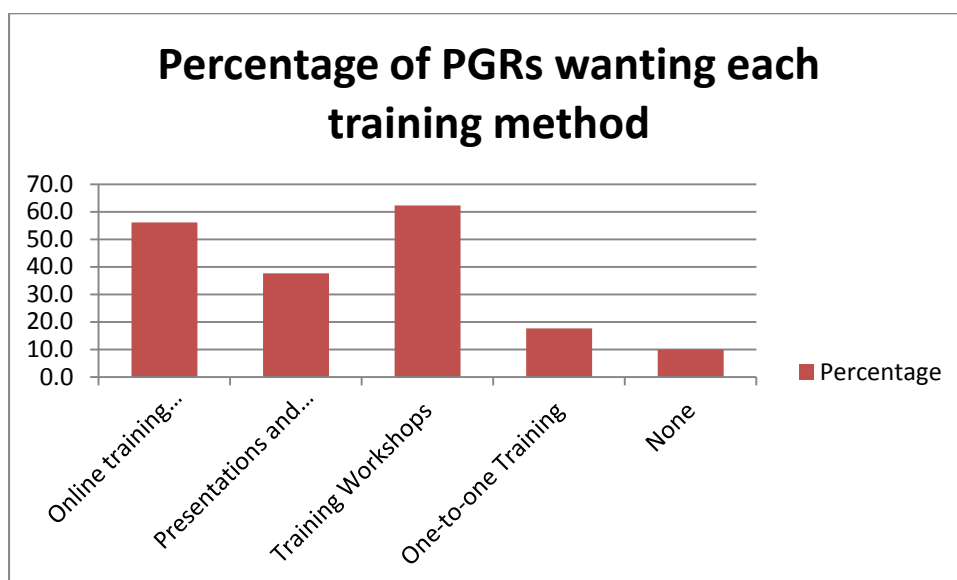


Chart 3.11.8: Graph showing the percentages of the format PGR students would like training to take

Again, differences can be seen between Colleges. For consistency, **Charts 3.11.9 and 3.11.10** show the results for CEMPS and CHUMS.

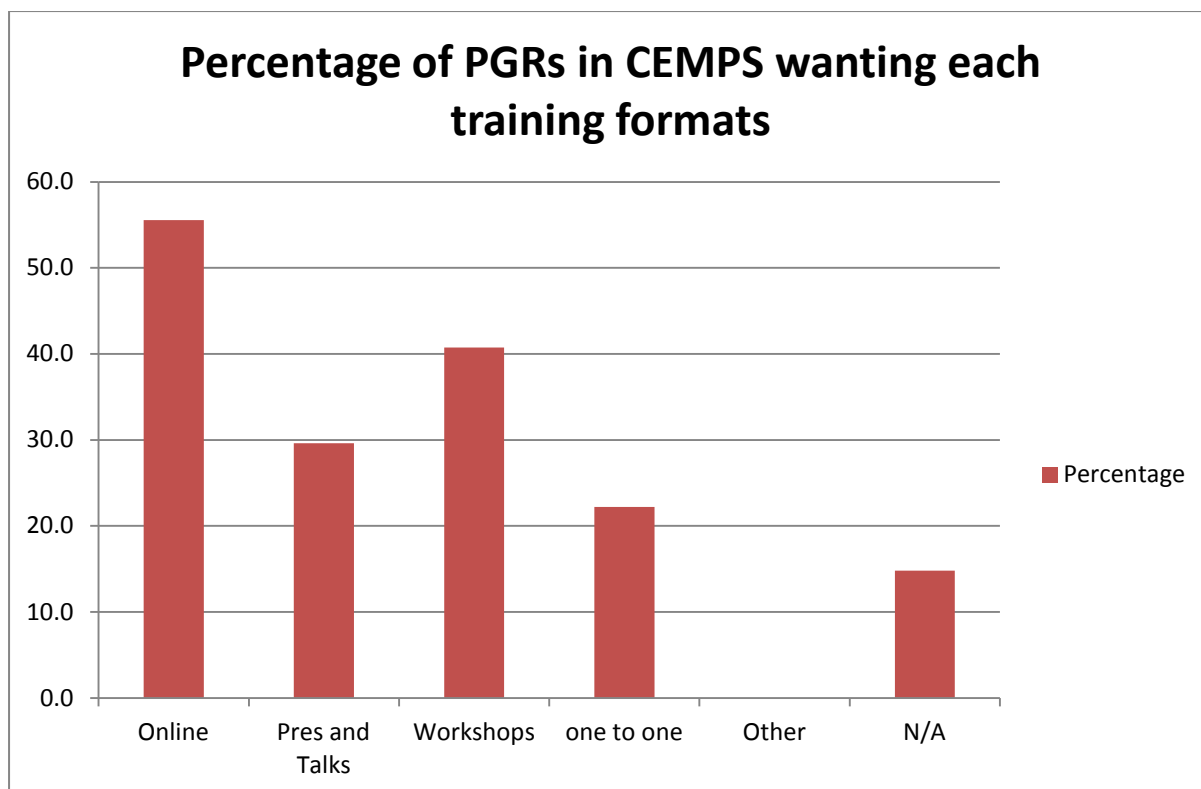


Chart 3.11.9: Graph showing the percentage of PGRs in CEMPS wanting each training format

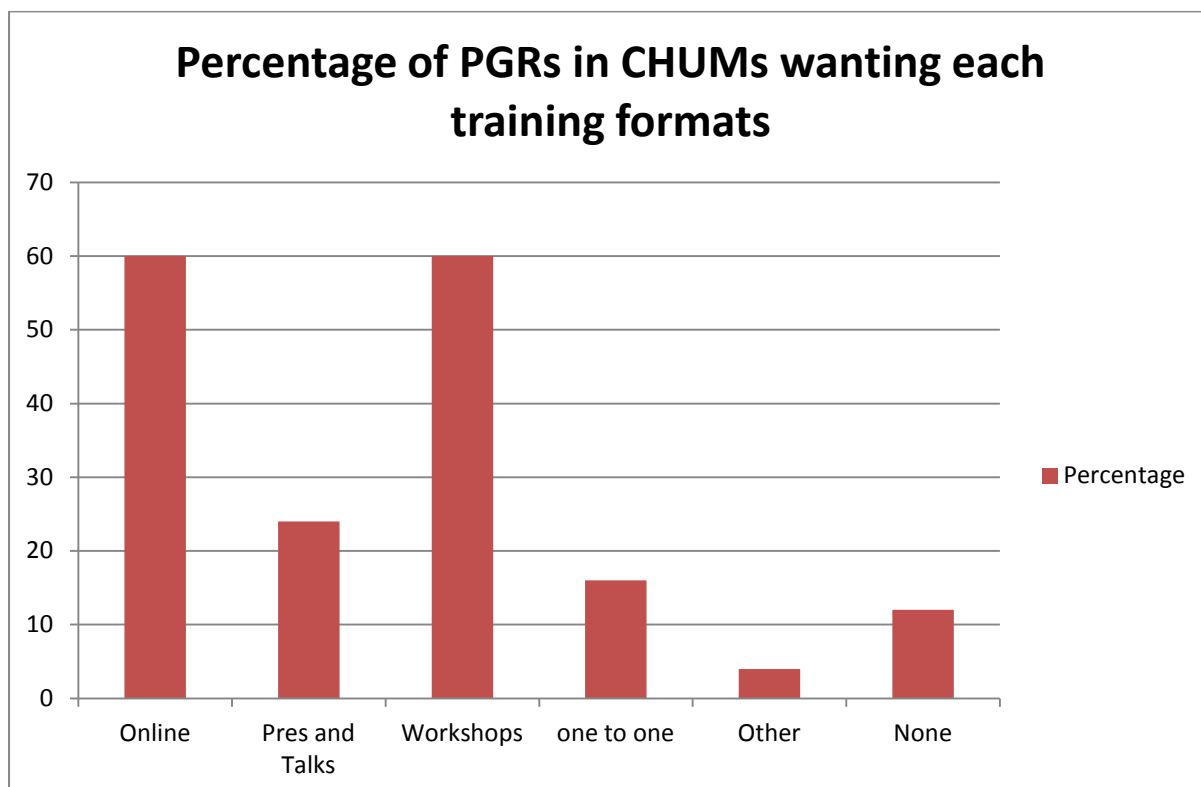


Chart 3.11.9: Graph showing the percentage of PGRs in CHUMS wanting each training format

An analysis of the results for training has shown that although general conclusions can be drawn (and some generic training will be an output of the Open Exeter project) there are specific needs and requirements within each College.

Interviewees have also stated that even some of the College-level training currently provided proves irrelevant to a number of the participants. Results show that one-to-one training is not always popular across Colleges and to all types of researcher, and that, apart from Business School researchers, online training materials and training workshops are the most popular ways of learning. Interviews have emphasised that training must include researchers based at St Luke's, Tremough, other University outposts such as research groups at the Knowledge Spa in Treliske Hospital in Truro, and those who are distance learners. Online materials will be especially important for these researchers but the benefits of workshop interaction and learning should not be denied to these distributed research teams.

3.12 Bibliographic Software

192 survey respondents use bibliographic software, almost 68%.

Broadly speaking, the more junior the researcher, the more likely they are to use a bibliographic software tool. Again, the figure for Associate Reader/Professor must be treated with some caution due to the small number of respondents (**Chart 3.12.1**).

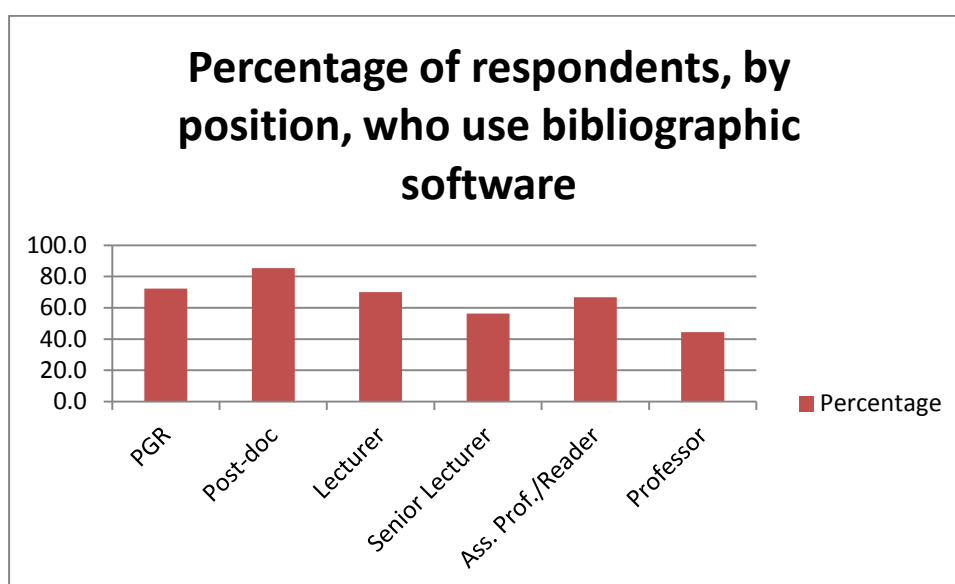


Chart 3.12.1: Percentage of respondents, by position held, who use Bibliographic Software

Interestingly, there also seems to be a split along HASS/STEM lines when it comes to bibliographic software (**Chart 3.12.2**). The primary bibliographic software package promoted and supported by training is Endnote, which costs approximately £70 to install on a private computer or laptop. Mendeley, which came second in the survey, has traditionally been seen as a tool for Science researchers although it is being increasingly taken up and used by other subject areas. Interviews and conversations at various events have highlighted the fact that many researchers are unaware of free bibliographic software such as Mendeley and Zotero but are very interested in trying alternatives to Endnote.

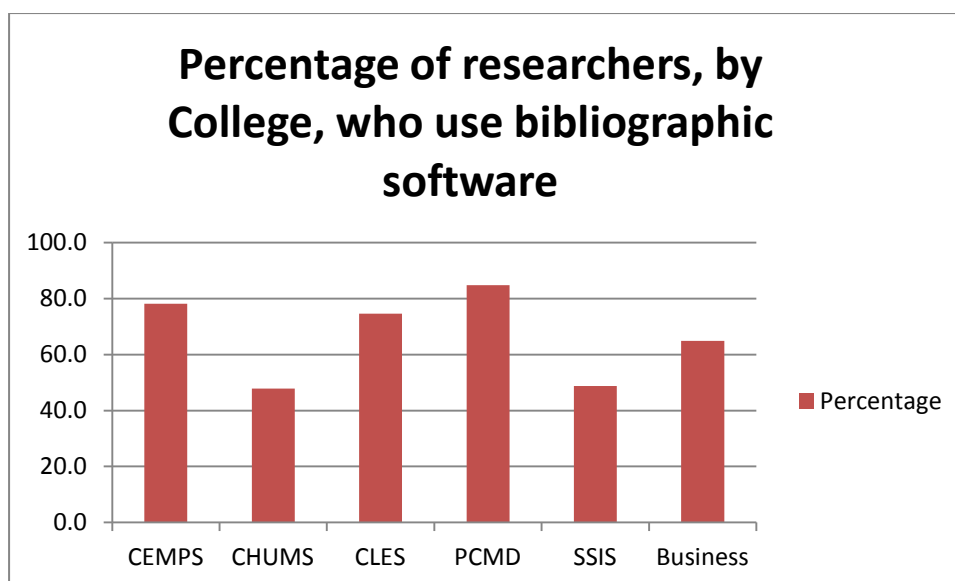


Chart 3.12.2: Percentage of researchers, by College, who use bibliographic software

Endnote was by far the most popular tool - 133 of the 192 respondents used Endnote (but as noted, Endnote is heavily promoted). Mendeley came a respectable second with 31 users. Of these 31, 14 are from CEMPS and 10 from CLES, three are from The Business School, two from CHUMS and two from the Medical School (PCMD). 15 respondents used Zotero, ten used BibTex/BibDesk, nine used Reference Manager; six used RefWorks, six used Papers, and four researchers used Jabref. One CHUMS PGR used Microsoft Word. A very small number of researchers used bespoke pieces of software.

The analysis of this section suggests that in future, training should be provided in more than one piece of bibliographic software. In addition, the encouragement of free to use software rather than paid software would probably increase take up and enable more efficient and better organised research.

Appendix 2: Open Exeter DAF Online Survey Questions

Available as a PDF from the Open Exeter [blog](#).

Appendix 3: Open Exeter DAF Interview Schedules and Consent Forms

Preliminaries (in all cases)

Sign consent form (in advance if possible). [Filming brief where applicable: Why he/she has agreed to be interviewed and why RDM is important.]

Data Protection – explain what will be done with data collected (in advance if possible).

Introduction to project and aims of the interview (covered in introductory email if possible).

Some basic definitions, e.g., what we mean by ‘data’ what is meant by ‘data management’ (sent in advance by email).

Open Exeter Interview Schedule – Academic/Researcher

About the interviewee and their research

Briefly describe your research.

Who are, or have been, the major funders of your research?

- Do funder requirements impact on your approaches to any aspect of data handling?

About their data

Some or all of these may need to be expanded on depending on responses to survey.

- What types of data do you typically or primarily work with?
- What are the main formats?
- What is the average size of a file (or dataset)?
- How much of it exists?
- Where is it held?
- How is it organised?
- Where is it created?
- How is it shared/transported?
- Is any of it sensitive?

Current and projected needs

Do University facilities meet all of your data (creating, using, sharing, storing, etc.) needs?

If no, explain why and what you would ideally like to improve?

Do you anticipate that your data handling needs will change in the next few years?

If yes, how?

- Storage space?
- Need to share more externally?
- Different styles of learning and delivering teaching?
-
- Different technologies and formats?

Data ownership

Does your project or group have a formal policy or guidelines on RDM?

- If yes – describe.

Are there any departmental or University policies or guidelines that affect any aspects of the life of your data?

Does anyone involved in your project or group have particular responsibility for the 'management' of data?

How, if you do, do you monitor data quality/authenticity?

Current vs completed data

How is your live (active or current) data distinguished from data that is no longer used?

At what point do you decide that data is no longer currently useful?

- What steps do you subsequently take?

How do you select what to keep and what to delete?

Open Access

What are your feelings about putting your research data on Open Access?

- Can you identify any immediate barriers to placing your work on OA?
- Have you ever made any of your work (data or published) on OA?
- Where?
- Why?

Would you put your research in an Exeter data repository?

- If no, why not?

Would it be useful to you if data could be cited in the same way as research papers?

Have you ever published your data as supplementary to a research paper?

- Where?
- How did you transfer the data to the publisher?

Data Management Planning

Have you ever written a Data Management Plan?

- What was your experience?
- Did you get help?
- From whom?

Supporting PGRs

What is your role in imparting RDM guidance to the PGRs you supervise?

Are you aware of PGRs' particular RDM needs or concerns?

What happens to PGRs' data if they leave?

- Is it kept?
- Where?
- Is it shareable?

Do you supply research data to PGRs?

- Where does this come from?

RDM training needs

Have you ever had any training in any aspect of RDM?

Would such training be useful for you, your project, group or PGRs?

How would you prefer to receive such training?

Research drive

Would a networked Research Drive be useful to you?

- How would you use it?
- What would be the benefits to you and your project?

Follow up

Would you be willing for us to base a case study on your project or research group?

Would you allow us to film you in the process of creating and managing your data?

Do you have any data that you would like to deposit in EDA?

- What is it?
- Size?
- Formats?
- Where is it?

Open Exeter Interview Schedule – College Administrator (Research)

About the interviewee and their role

What is your role within your College and what does it involve?

Are there any limitations on what you are able to achieve in your role, e.g., budgetary, staff resources?

Communication

What College, departmental and University meetings, boards or committees are you involved with?

How is research being carried out in College disseminated (both in the College and around the University)?

How do the various research groups and projects within your College communicate?

Is there a general awareness of funders' requirements regarding research data and OA?

Whose responsibility is it to make researchers aware of internal and external factors affecting their research?

Policy

Who are the major funders of research in your College?

Does your College or any department or group within the College have a formal policy or guidelines on RDM that you are aware of?

- If yes, describe.

Are there any College, departmental or University policies or guidelines that affect any aspects of RDM? For example:

- Ethical
- Code of Practice in Research

Does anyone involved in your College have particular responsibility for research data?

Open Access

What are your feelings about putting College research data on Open Access?

- Can you identify any immediate barriers to placing College work on OA?
- Have you ever made any of your own work (data or published) on OA?
- Where?
- Why?

Would you put your own or College research in an Exeter data repository?

- If no, why not?

Current and projected needs

Do University facilities meet all of your College needs regarding RDM?

- If no – explain why and what you would ideally like to improve?

Do you anticipate that College technology/data needs will change in the next few years?

- If yes, how?

RDM training needs

Have you ever had any training in any aspect of RDM?

Would such training be useful for you and staff and students in your College?

How would you prefer to receive such training?

Research drive

Would a networked Research Drive be useful to your College?

- How would College members use it?
- What would be the benefits to you and your College?

Follow up

Would you allow us to attend any College meetings to promote the work of the Project?

Would you be willing for us to base a case study on a project or research group within your College?

Are you aware of anyone in your College who has any data that could be deposited in EDA?

- What is it?
- Size?
- Formats?
- Where is it?

Open Exeter Interview Schedule – IT and Technical Support Staff

About the interviewee

What is your role within your department?

Issues and queries

What is the main type of support you provide?

Who/which group of people are your biggest customers?

What are the main queries you are asked to help with?

Do particular groups have specific queries?

Are there any queries that come up repeatedly?

Are there any queries common to all staff and students?

Are there any barriers to or restrictions on the kind of support you are able to provide?

About the data they support

What types of data do you typically come across?

What are the main formats?

What is the average size of a file (or dataset)?

Where is it held?

Is any of it sensitive?

Current and projected needs

Do University facilities meet all of your needs regarding providing support?

- If no, explain why and what you would ideally like to improve?

Do you anticipate that these needs will change in the next few years?

- If yes, how?

Policy

Does your department or the University have any formal policy or guidelines on data that affect how you work?

- If yes, describe.

Teaching & Learning

Do you provide any formal training sessions in your department?

- What sort?
- Face to face
- Online
- Written

How have you learnt any RDM skills that you regularly use?

Would peer to peer sharing of RDM knowledge be a good way to pass on skills?

Open Access

What are your feelings about putting departmental research data on Open Access?

- Can you identify any immediate barriers or issues?

Could you see yourself helping people to put their research in an Exeter data repository?

- If no, why not?

RDM training needs

Have you ever had any training in any aspect of RDM?

Would such training be useful for you?

How would you prefer to receive such training?

Research drive

Would a networked Research Drive be useful to you?

- How would you use it?
- What would be the benefits to you and your project?

Follow up

Can you suggest other people to interview or suggest a likely project or group for a case study?

Would you allow us to film you in the process of providing RDM support?

Are you aware of anyone who has any data that could be deposited in EDA?

- What is it?
- Size?
- Formats?
- Where is it?

Open Exeter Interview Schedule – Postgraduate and Post Doctoral

About the interviewee and their research

Briefly describe your research.

Who are, or have been, the major funders of your research?

- Do funder requirements impact on your approaches to any aspect of data handling?

About their data

Some or all of these may need to be expanded on depending on responses to survey.

What types of data do you typically or primarily work with?

What are the main formats?

What is the average size of a file (or dataset)?

How much of it exists?

Where is it held?

How is it organised?

Where is it created?

How is it shared/transported?

Is any of it sensitive?

Current and projected needs

Do University facilities meet all of your data (creating, using, sharing, storing, etc.) needs?

If no, explain why and what you would ideally like to improve?

Do you anticipate that your data handling needs will change in the next few years?

If yes, how?

- Storage space?
- Need to share more externally?
- Different styles of learning and delivering teaching?
- Different technologies and formats?

Data ownership

Does your project or group have a formal policy or guidelines on RDM?

- If yes, describe.

Are there any departmental or University policies or guidelines that affect any aspects of the life of your data?

Does anyone involved in your project or group have particular responsibility for the 'management' of data?

How do you monitor data quality/authenticity?

Current vs completed data

How is your live (active or current) data distinguished from data that is no longer used?

At what point do you decide that data is no longer currently useful?

- What steps do you subsequently take?

How do you select what to keep and what to delete?

Open Access

What are your feelings about putting your research data on Open Access?

- Can you identify any immediate barriers to placing your work on OA?
- Have you ever made any of your work (data or published) on OA?
- Where?
- Why?

Would you put your research in an Exeter data repository?

- If no, why not?

Data publication

Would it be useful to you if data could be cited in the same way as research papers?

Have you ever published your data as supplementary to a research paper?

- Where?
- How did you transfer the data to the publisher?

Data Management Planning

Have you ever written a Data Management Plan?

- What was your experience?
- Did you get help?
- From whom?

Teaching & Learning

If you teach, what is your role in imparting RDM guidance to students?

How have you learnt any RDM skills that you regularly use?

What was your experience of being supervised in relation to RDM support?

Would peer to peer sharing of RDM knowledge be a good way to pass on skills?

What would happen to your data if you left?

- Could I take it with me?
- Would it be kept?
- For how long?
- Where?
- Would it be shareable?

RDM training needs

Have you ever had any training in any aspect of RDM?

Would such training be useful for you, your project, or research group?

How would you prefer to receive such training?

Research drive

Would a networked Research Drive be useful to you?

- How would you use it?
- What would be the benefits to you and your project?

Follow up

Would you be willing for us to base a case study on your project or research group?

Would you allow us to film you in the process of creating and managing your data?

Do you have any data that you would like to deposit in EDA?

- What is it?
- Size?
- Formats?
- Where is it?



Photograph, Filming and Sound Recording Consent

I give permission to the JISC Open Exeter Project and the University of Exeter for the use of photographs and/or film and/or sound recording and, if applicable, written profile and/or agreed quotes to be used in print and electronic publications and promotional material, including but not limited to, digital reproduction for use on the web, and other media as appropriate.

I do not object to the Open Exeter Project and the University storing copies of the photograph(s) and/or video/film and/or sound recording in order to further the work of the Project.

Information will be used only in accordance with the Data Protection Act 1998.

Signed

Print name

Title

College/Department

Email

Date





Photograph, Filming and Sound Recording: Verbal Consent Form

This form records the oral consent of persons in a group setting, e.g., large group (whose image is featured in conventional and digital photographs and/or analogue or digital film and whose spoken word is recorded in a separate or accompanying sound recording) where it is impractical to obtain individual consents from all persons present.

Event/Activity

Date of Event/Activity

On behalf of the Open Exeter Project and the University of Exeter, I confirm that identifiable individuals in the photograph(s) and/or film(s) and/or sound recording(s) taken at the above event or activity have consented to being photographed, filmed and if applicable, their spoken word recorded, and have consented to the media recording being made available on the Internet via University and other organisations' websites.

Details of Open Exeter or University representative gaining consent of the group:

Signed

Print name

Title

College/Department

Email

Date

JISC