Research Data Management

Analysis of Research Data Management Survey

September 2013

16 October 2013

Introduction

A Research Data Management (RDM) survey was launched in September 2013 and disseminated to research active staff within the College, via email distribution lists and other networks. A copy of the survey disseminated is available as Appendix A.

The survey was disseminated to approx. 350 research staff and we received 80 complete responses (8 incomplete). This constitutes a 23% response rate for the survey. This analysis is based on the 80 complete responses received.

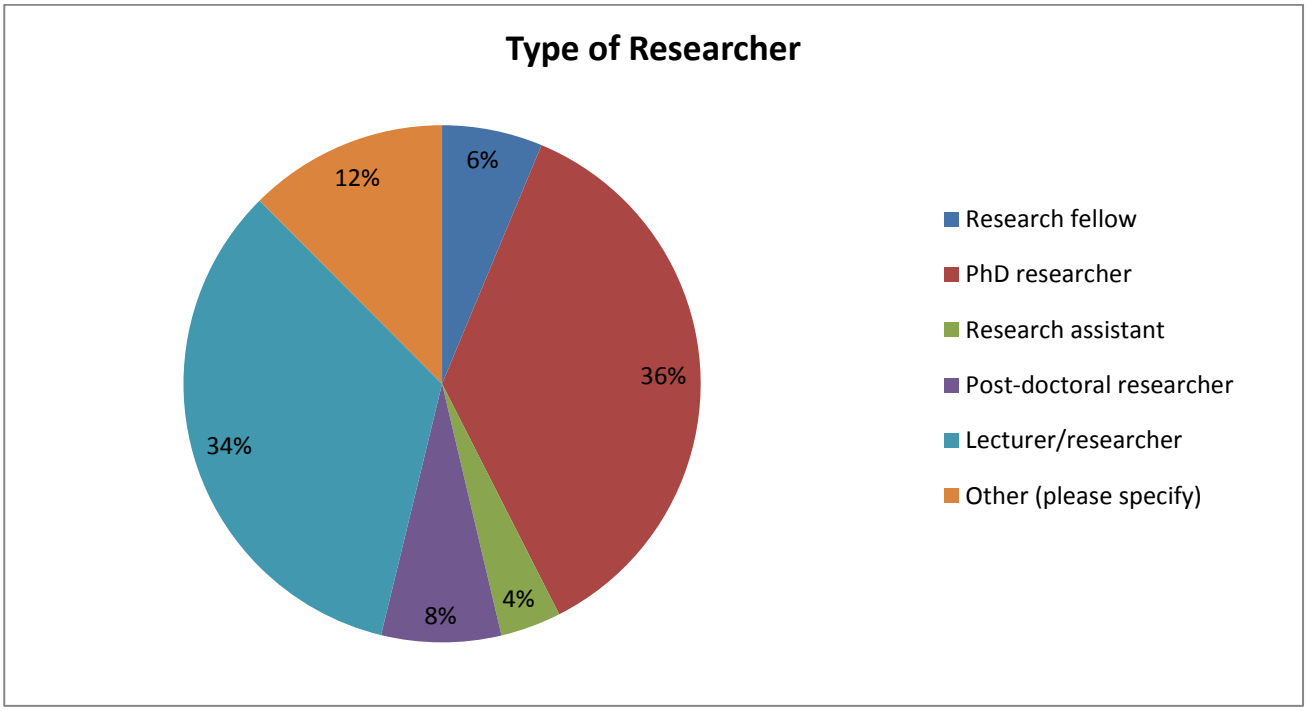
Summary

- Generally, the knowledge and understanding of RDM practices and processes is low.
    - 29% create RDM plans
    - 32% are aware of RDM requirements from funders
    - 30% currently make research data available via open access / repositories
    - 29% document appropriate metadata for research data.
- Only 10% of respondents have received any form of RDM staff development and most respondents would welcome staff development on all aspects of Research Data Management.
- Currently data is stored in a variety of locations, with locally stored data (e.g. portable hard drives, computer hard drives and project specific servers) accounting for a significant percentage of storage. This data may not be backed up frequently or securely so constitutes a risk which needs to be evaluated and mitigated.
- Data backup is undertaken in some form (either daily, weekly or ad-hoc) by 95% of respondents. This backup is mainly undertaken using external media (portable hard drive, memory stick etc). Knowledge of college backup processes is 54%.
- A variety of tools and data formats are used in the creation of research data. Whilst Microsoft Excel is the most commonly used tool (68%) a range of other tools and formats are used. This information can be used to create tool specific support and guidance on open standards and metadata requirements.
- The majority of respondents (63%) have storage requirements of less than a Terabyte. 14% have requirements up to 10Tbs and 7% have requirements of over 10 Terabytes.
- 17% of respondents stated they were unaware of the storage requirements they have. This indicates a key staff development need to support Research staff in their understanding of data size and storage issues. However, it is likely a number of these will be on the lower end of storage requirements.
- A significant level of research data is created/documented using paper-based methods. 73% of respondents use paper as part of their research. The appropriate storage and archiving of paper records remains a college-wide concern.

1. General Information

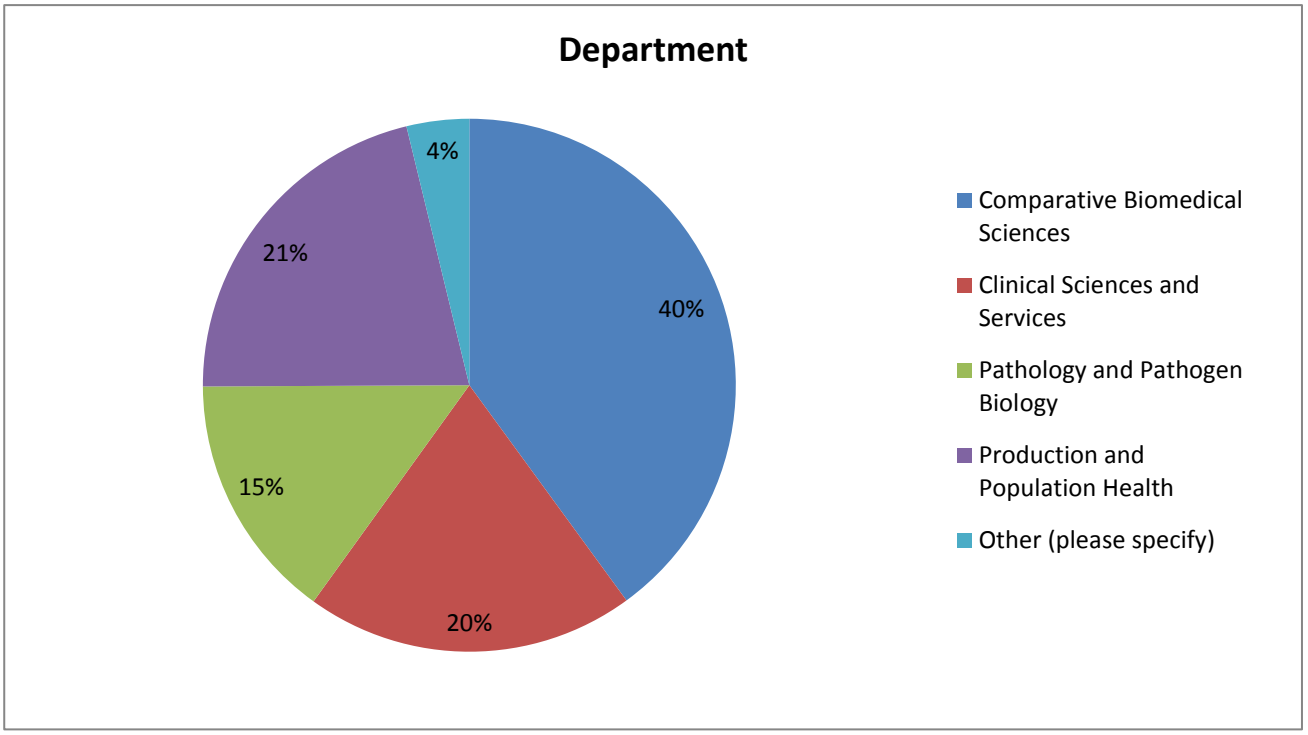
1.1. Please state which type of researcher you are:

The majority of responses came from those who class themselves as either PhD Researchers (36%) or Lecturer/Researcher.



1.2. Please list your department:

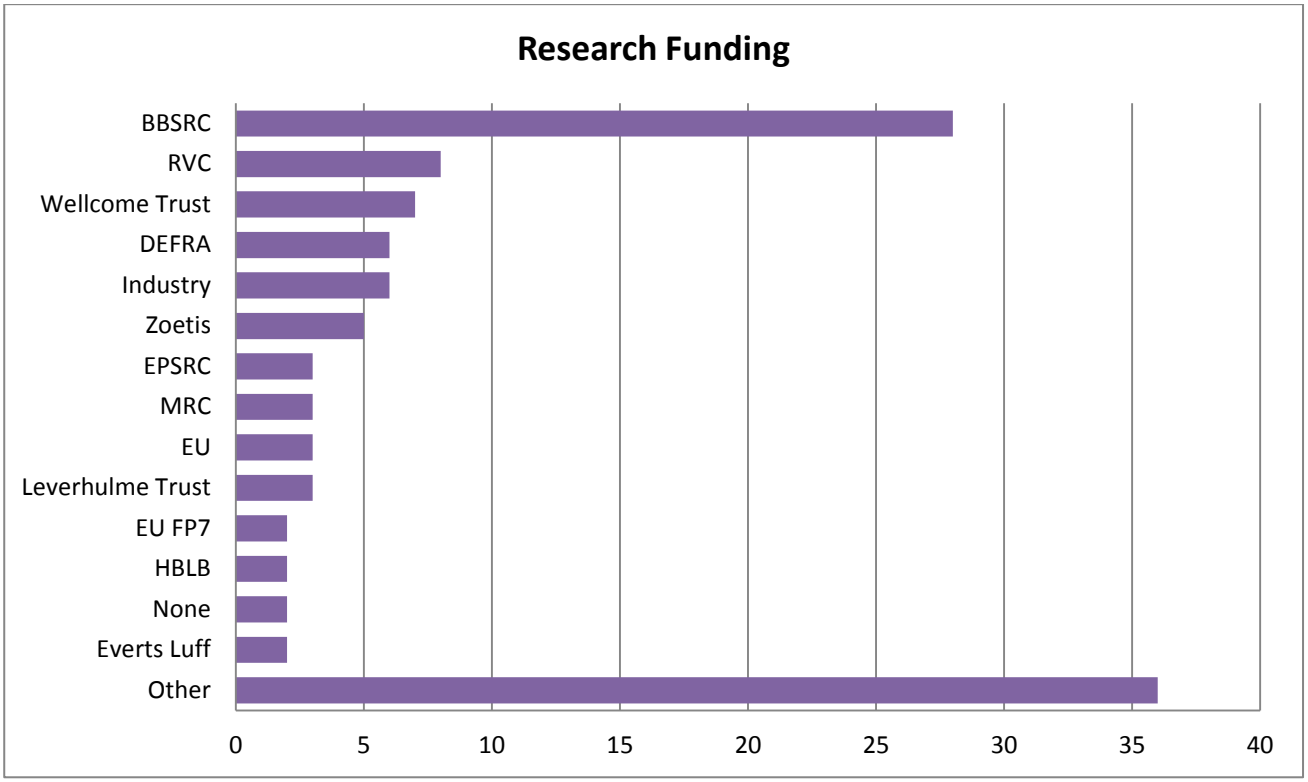
The majority of responses (40%) came from those within Comparative Biomedical Sciences. Responses were relatively evenly distributed between the three other departments.



1.3. Who funds the research project(s) you are currently undertaking?

Respondents were asked to list their top three funders for research. As expected the BBSRC is the most common funding provider.

The other category contains any funders who were only listed once. These could be analysed further to show the rough categories in which they sit, but generally a number of these would appear to be in the 'Industry' category.



Other

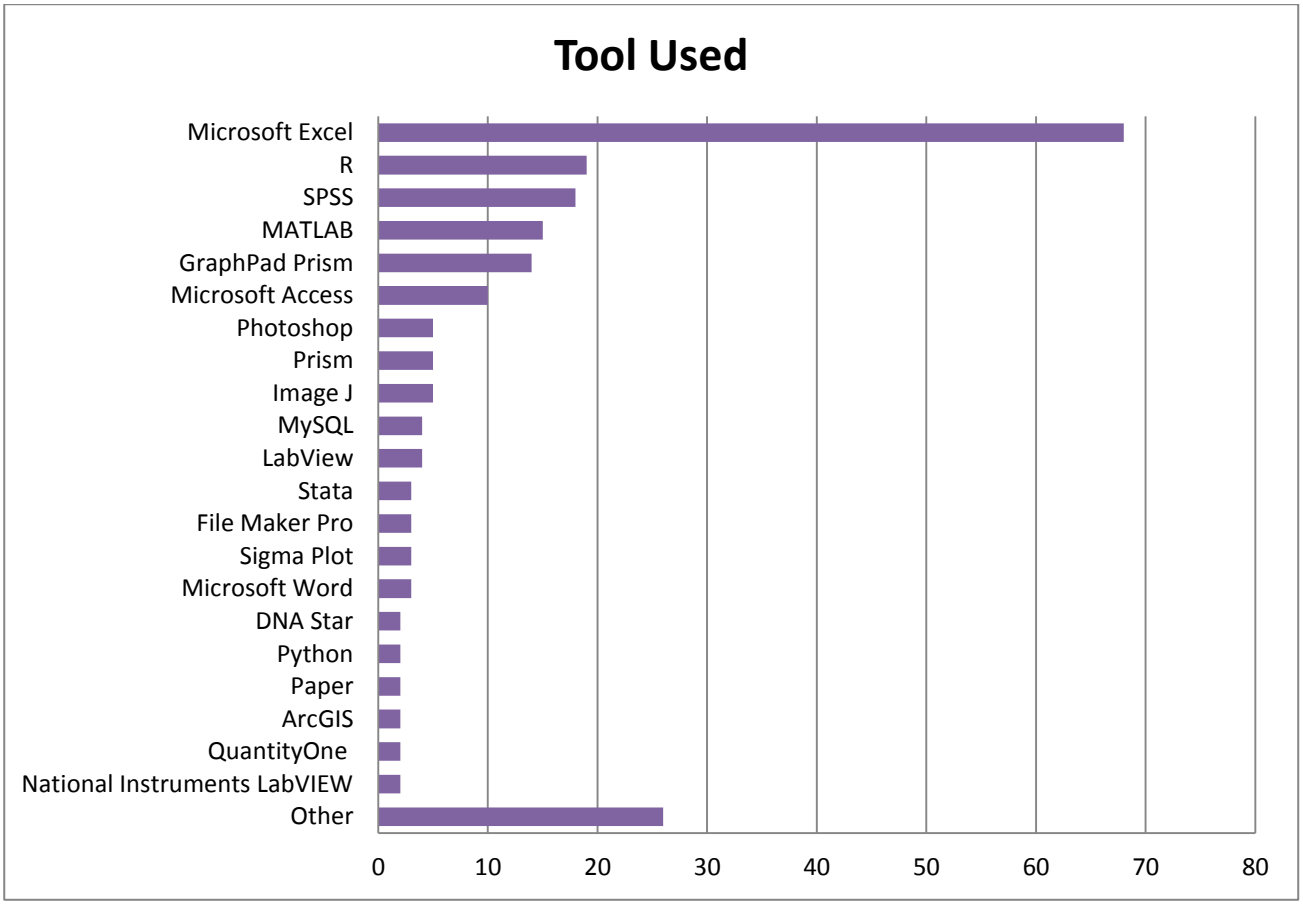
- Agrotec
- ASFORCE project
- Boehringer Ingelheim
- Breed Societies
- CSS
- DairyCo
- DFID
- Diabetes UK
- DRF
- Dstl
- EBLEX
- ERC
- European Commission
- European Society of Veterinary Dermatology
- FAO
- FAZD
- FCC
- FFI
- HFSP Human Frontier Science Program
- Horserace Betting Levy Board
- Humane Slaughter Association
- KCCT
- Marie Curie
- Medel
- Merck Animal Health
- Muscular Dystrophy Association
- Muscular Dystrophy Campaign
- Norbrook Pharmaceuticals
- Novartis
- PPCT
- Protexin
- Royal Canin
- Royal Society Petplan
- Senexis
- SRUC
- Waltham

2. Your Research Data

2.1 Please list the three most common tools you use to create and/or manipulate your research data.

In order to identify the tools being used throughout the college to work with research data, users were asked to list the top 3 packages used. Microsoft Excel is the most commonly used tool throughout the college, with other statistical analysis packages (e.g. R, SPSS and Microsoft Access) also commonly used.

The chart below lists any tool used by 3 or more respondents, those with one or two responses are combined and listed within 'Other'.



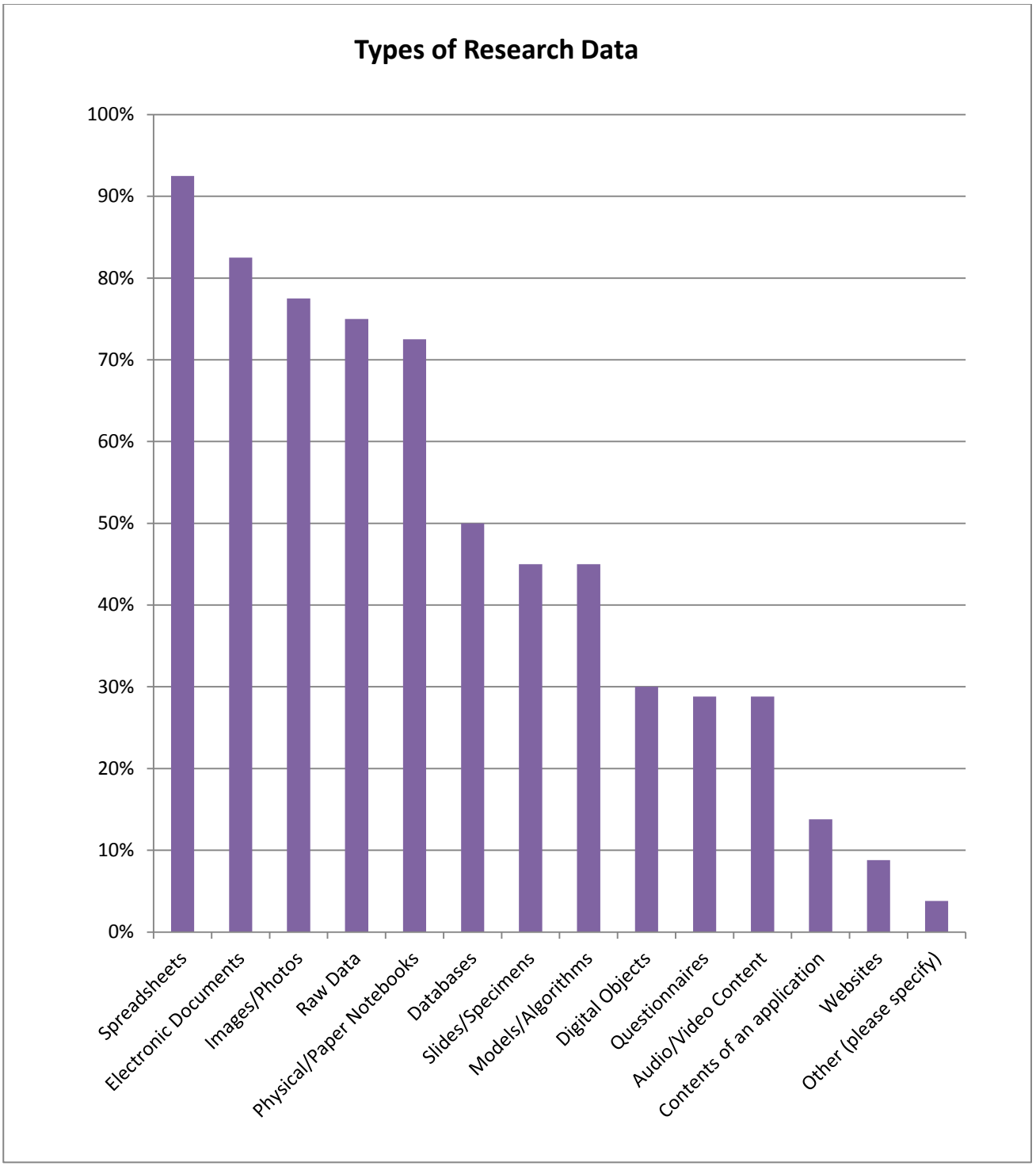
Other

- ASReml
- CLC Main Workbench
- custom C++/ Python code
- Custom software/code
- Dassault Systemes Abaqus
- Etholog
- GeneSpring
- Image processing
- IPA
- LabChart
- LaVision
- Mathematica
- Maya
- MEGA5.1
- Microsoft PowePoint
- Nvivo
- Origin (OriginLab)
- Paraview
- PLINK
- SAS
- SIMM
- Solidworks
- Stat
- Statistic package
- Unistat
- Video

2.2 What types of research data are created or worked with during the life cycle of the project?

As the use of tools would indicate, the most commonly created types of research data are Spreadsheets and electronic documents. The spread of responses over a variety of tools provides evidence for the staff development/guidance required to support users in the creation of open source formats and metadata for datasets.

Also of note is the fact 73% of researchers use Physical/Paper Notebooks as part of their research. The physical storage and archiving of paper resources still requires a college-wide solution.



2.3 Where do you store your data during the life cycle of your project(s)?

Respondents were asked to note where they store research data during the lifecycle of the project. As the list below shows, there are a variety of locations where data is stored. Storage can be roughly divided into four categories:

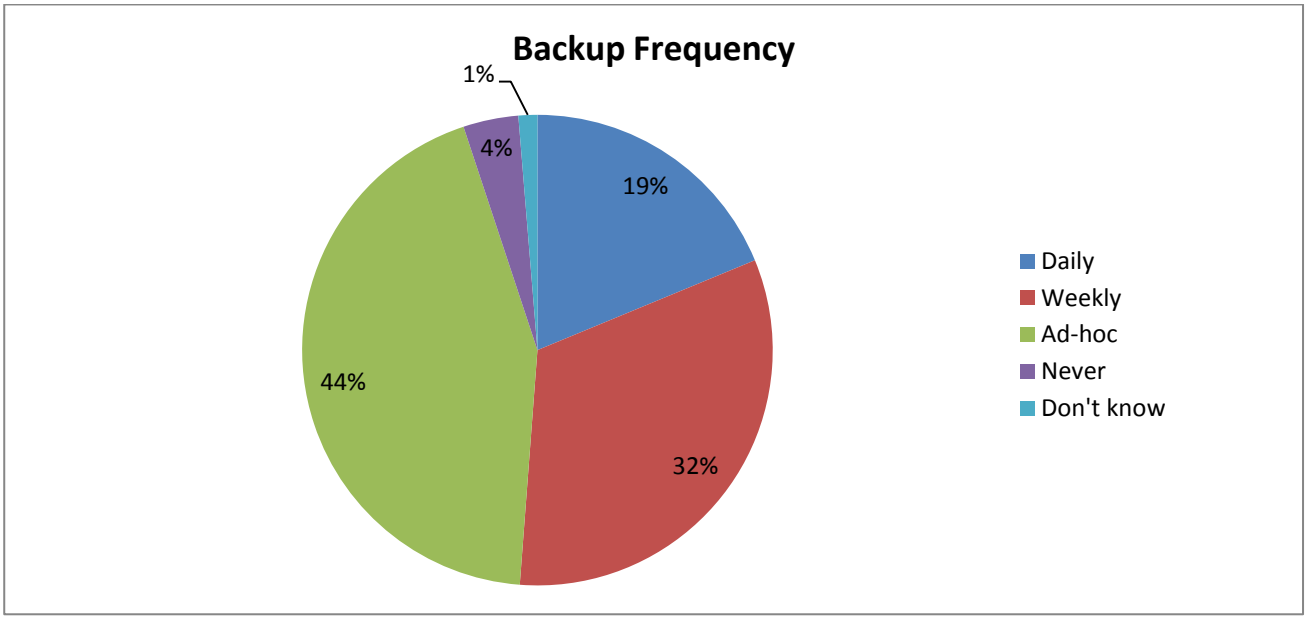
- **Stored locally** (portable storage, hard drives / local drives etc). These sources would traditionally not have automatic back-up processes in place.
- **Stored centrally** (on the H:drive, R:drive or O:drive etc). These sources would be backed up routinely by the Colleges backup process or cloud services.
- **Stored physically** (on paper, photographs, slides etc). As physical entitles, there are unlikely to have a backup process.
- **Other**

Of most concern are those users who store their data locally as they may not be undertaking a robust backup process for their data, resulting in a high risk for the College for the security of data.

Answer Options	Percent	Count
Stored Locally		
Portable storage device e.g. external hard drive, USB stick	85%	68
Hard disk/local drive on College computer or laptop	63%	50
Hard disk/local drive on personal computer or laptop	40%	32
Instrument which generates data	29%	23
Project or user specific server e.g. NAS Box, PC shared drive	18%	14
CD/DVD	13%	10
Stored Centrally		
Your College supplied personal H:drive	56%	45
A College shared drive e.g. R:drive or O:drive	29%	23
Web based service e.g. Dropbox, Google Drive	44%	35
Email system	33%	26
Via a service provided by a collaborating institution	8%	6
Stored Physically		
On Paper	51%	41
Slides	15%	12
Physical Photographs	6%	5
VHS/Video Cassette	0%	0
Floppy Disk/ZIP Drive	0%	0
Microfiche	0%	0
Other		
Other (please specify)	4%	3

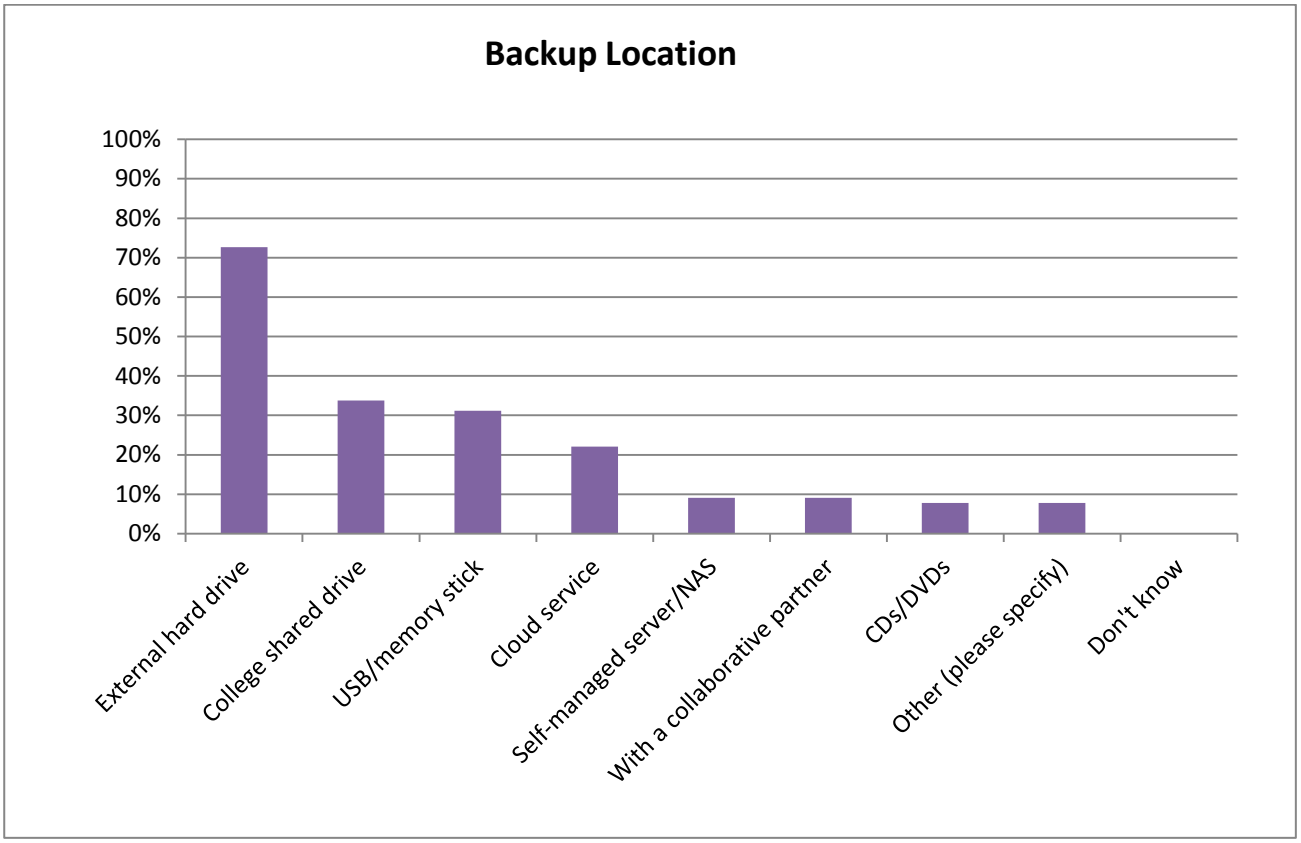
2.4 How frequently do you back up data created during the life cycle of your project(s)?

Respondents were asked to note how frequently they backed up data. 95% of respondent had some form of backup process in place, but the majority of these (44%) were on an ad-hoc basis. Whilst only 5% of respondents stated they did not backup data or did not know of a backup process, its important the location of this data is identified and any potential risks mitigated.



2.5 If you do back up your data, where do you back it up to?

Respondents were asked to note where they backed up their data. Most responses included some form of external media e.g. External hard drives (73%) and USB/memory sticks (31%).

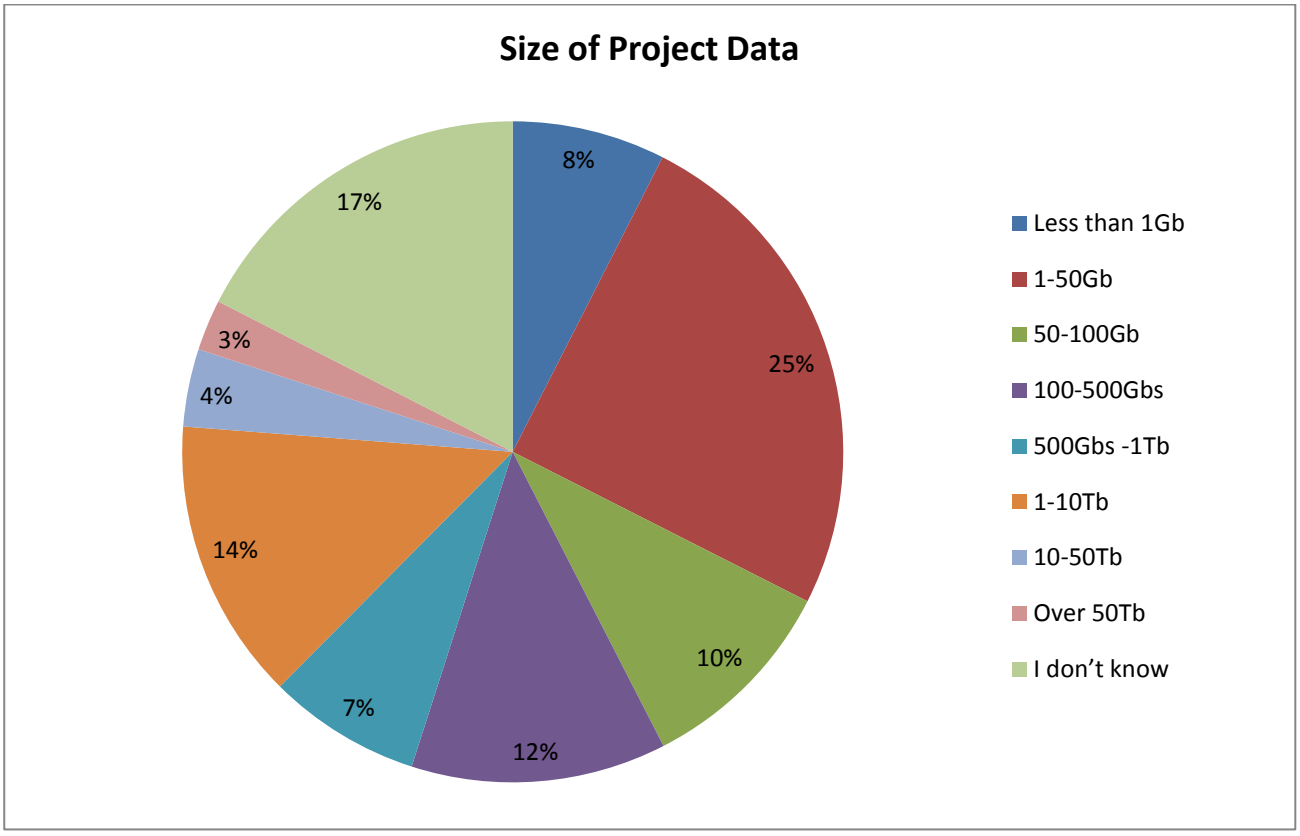


2.6 Approximately how much data do you expect your project(s) to create during its life cycle?

Respondents were asked to note approximately how much data they expect their research projects to generate. This information can be used to support the creation of a costing model for Research Data Management.

The majority of respondents (63%) have storage requirements of less than a Terabyte. 14% have requirements up to 10Tbs and 7% have requirements of over 10 Terabytes.

17% of respondents stated they were unaware of the storage requirements they have. This indicates a key staff development need to support Research staff in their understanding of data size and storage issues. However, it is likely a number of these will be on the lower end of storage requirements.



2.7 Are you aware of the backup process the College undertakes for data stored on its servers?

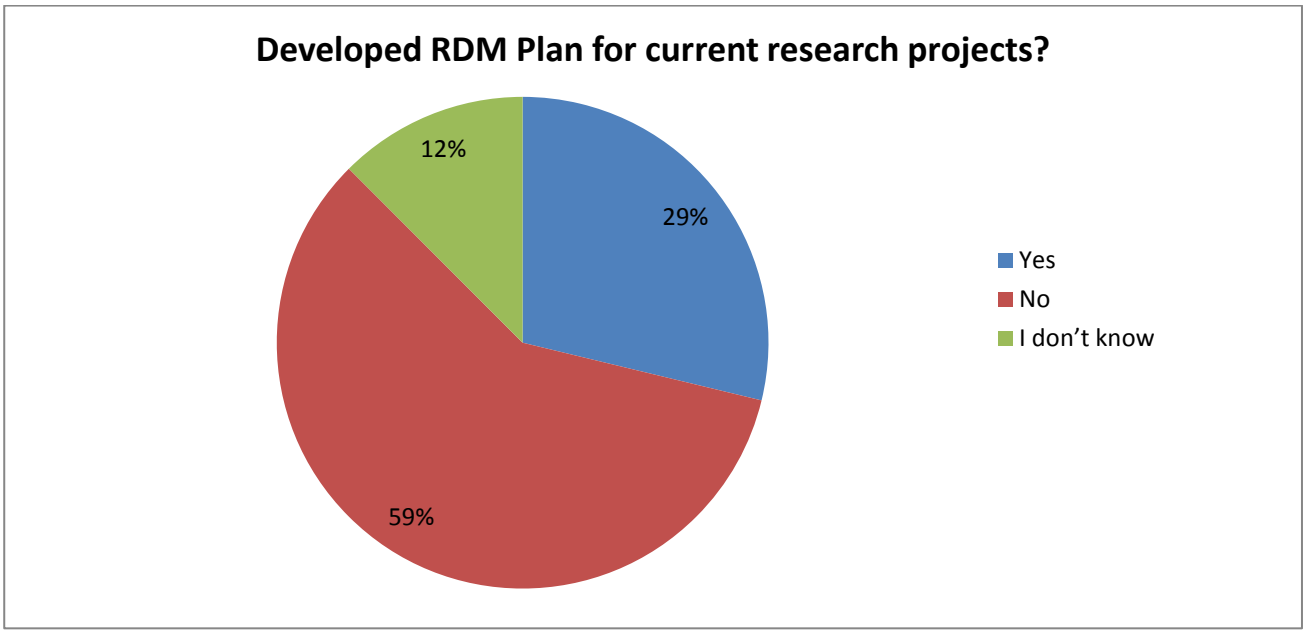
Respondents were asked to state whether they were aware of the backup processes the College undertakes for data. Just over half (54%) were aware of the processes. Again, this indicates a focus for staff development, ensuring college wide understanding of processes.

Answer Options	Percent	Count
Yes	54%	43
No	46%	37

3. Research Data Management

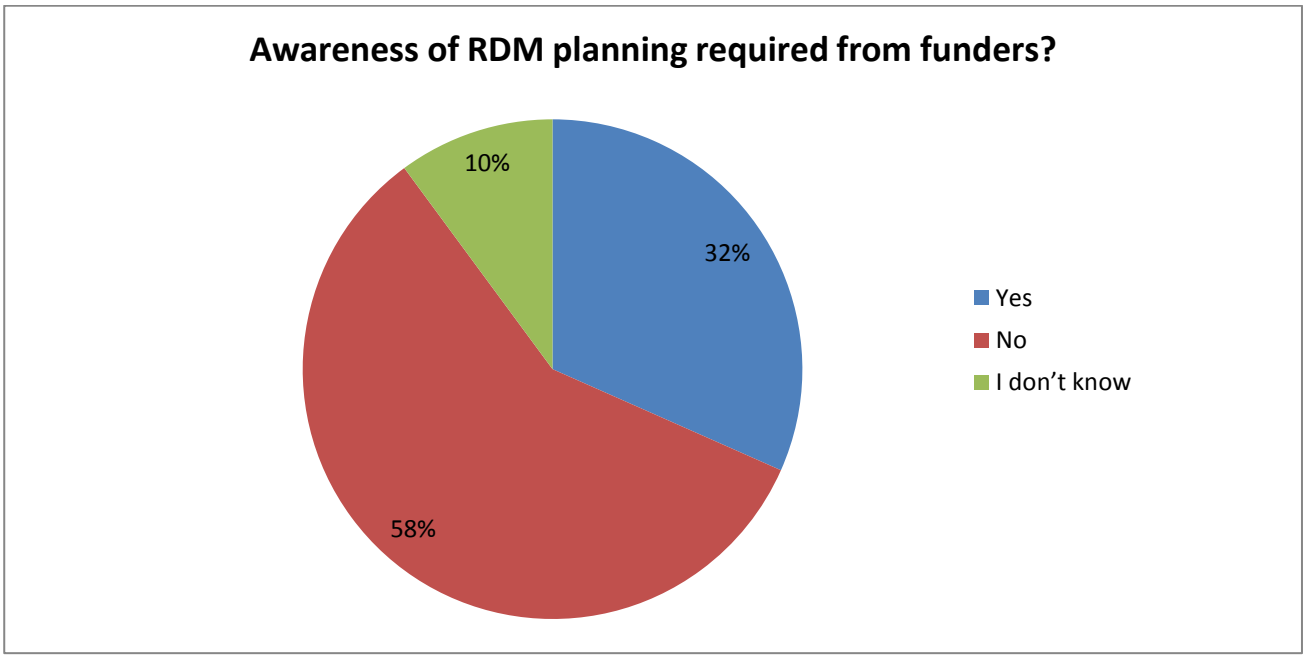
3.1 Have you developed a Research Data Management Plan for any current research projects?

71% of respondents have either not created an RDM plan or are unaware of whether RDM plans have been created for current research projects. This indicates a key element and requirement for staff development.



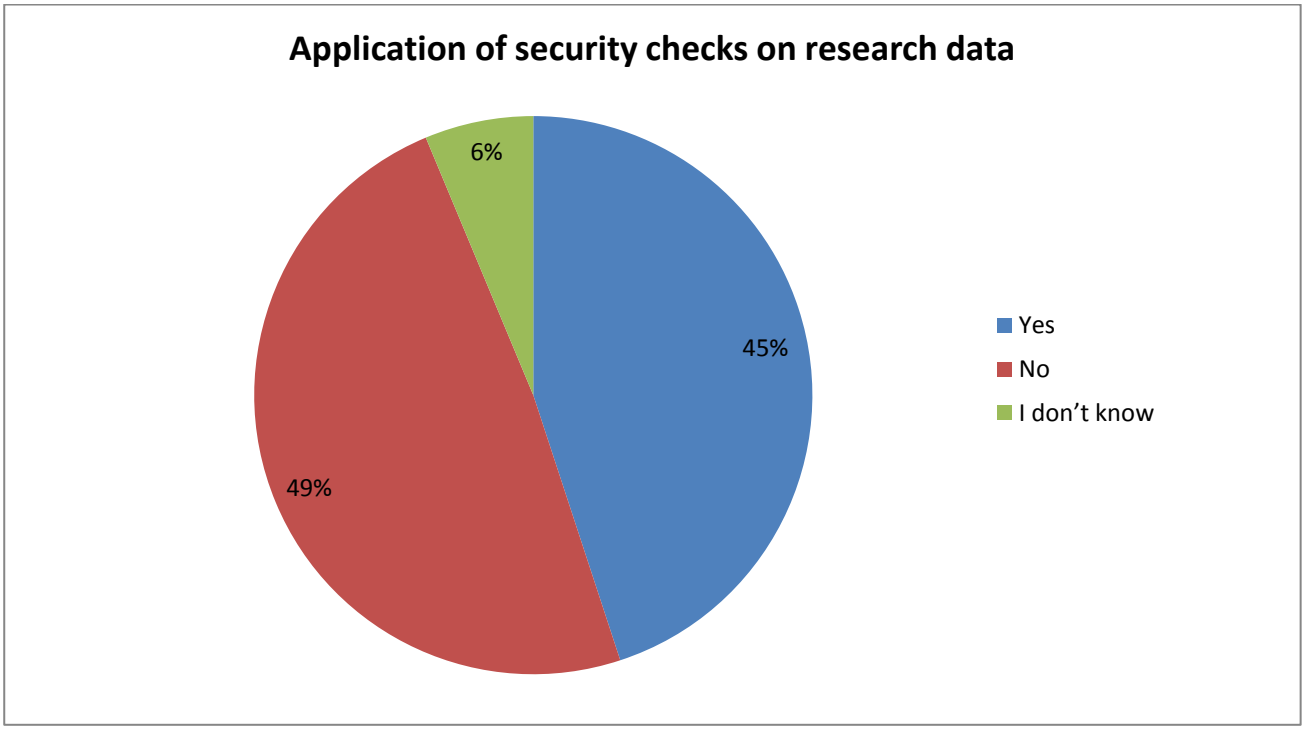
3.2 Are you aware of any specific research data management planning required from your funders?

Roughly aligned with the responses above, the awareness of RDM requirements of funders is only 32%. Specific and targeted staff development is required to ensure Researchers are aware of their requirements and how to fulfil them.



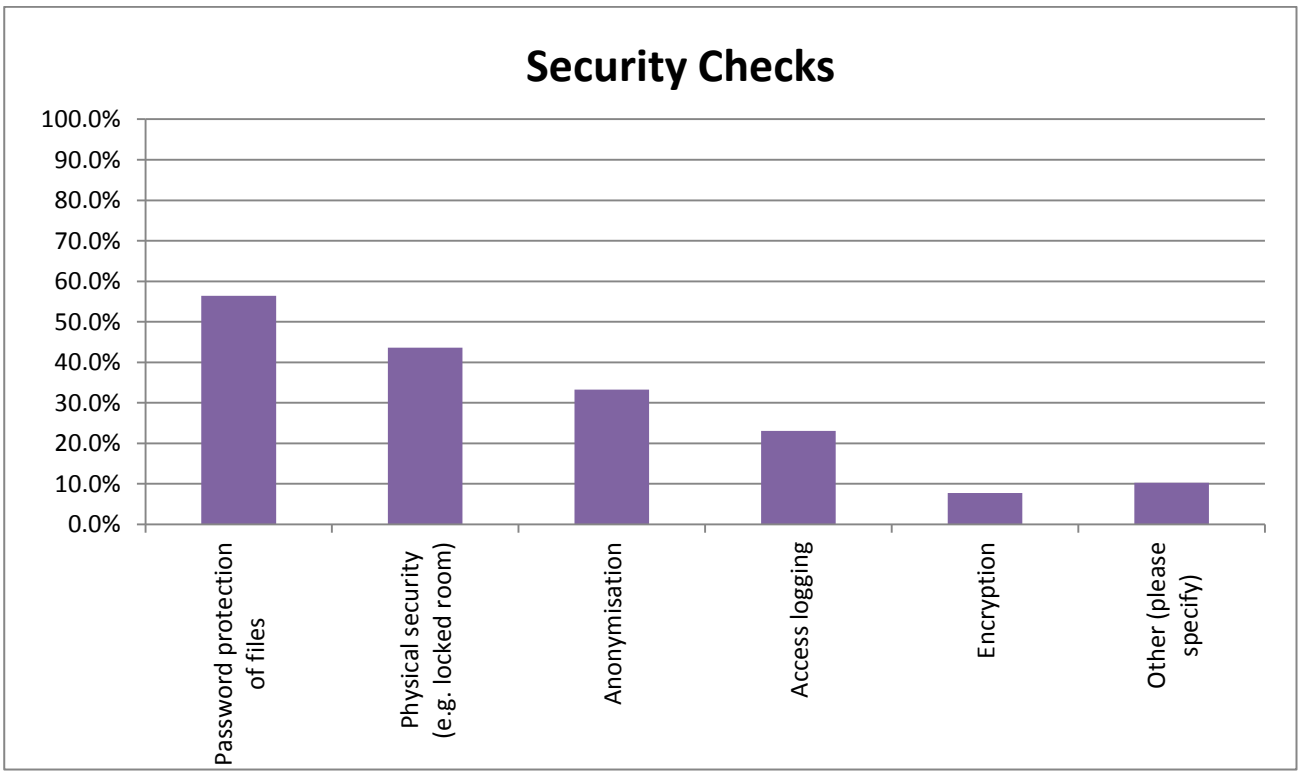
3.3 Do you currently apply security checks on some or all of your research data?

Just over half of respondents (55%) do not apply security checks to their data or are unaware of any security used. Data security is a key requirement to be built into a staff development programme.



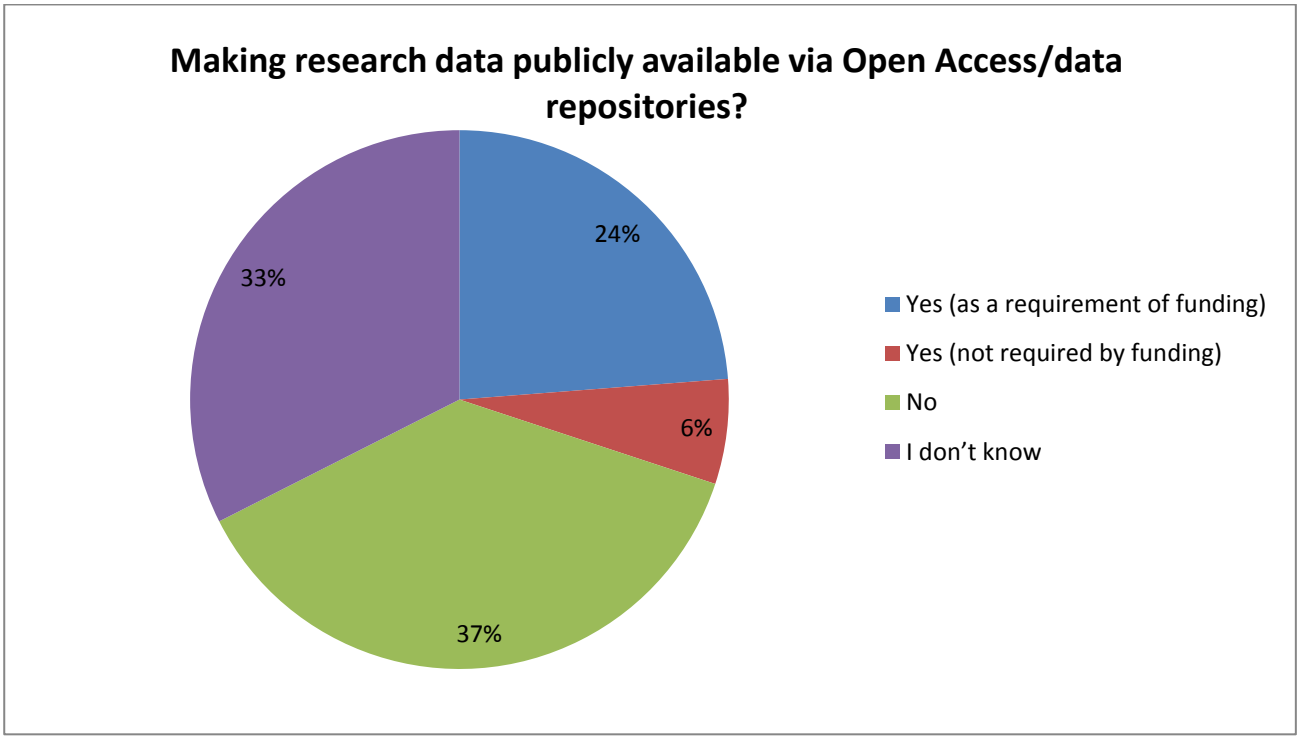
3.4 If yes, please detail any security checks you place on your data.

Of those who do apply security checks to their data, a variety of methods are used. Of interest is how much data is physically stored for security (44%).



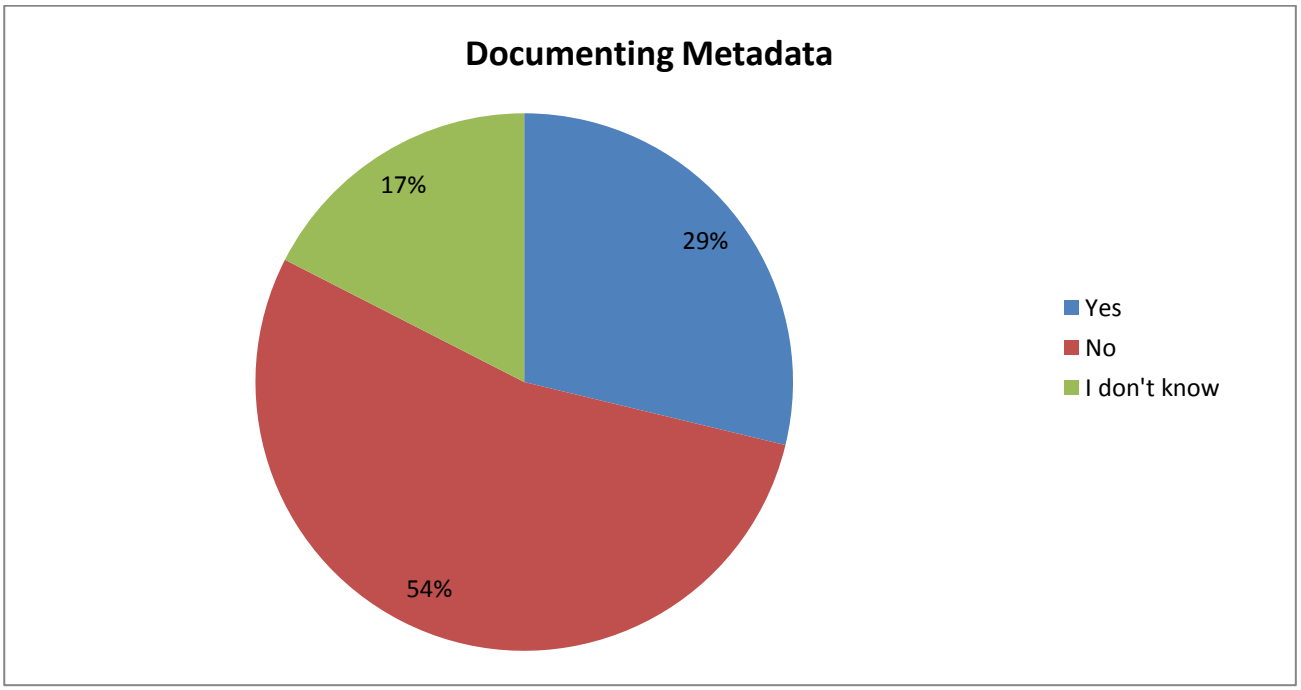
3.5 At the end of a project, do you make your research data publicly available via Open Access/data repositories?

Only 30% of respondents are currently making their research data publically available. Mechanisms need to be available to support researchers in ensuring they understand how and when to make data publically available.



3.6 Do you document or record any metadata about your data, to make it more meaningful to others?

54% of respondents are currently not documenting metadata about data sets. In order to support research staff, targeted staff development will be required.



4. Staff Development

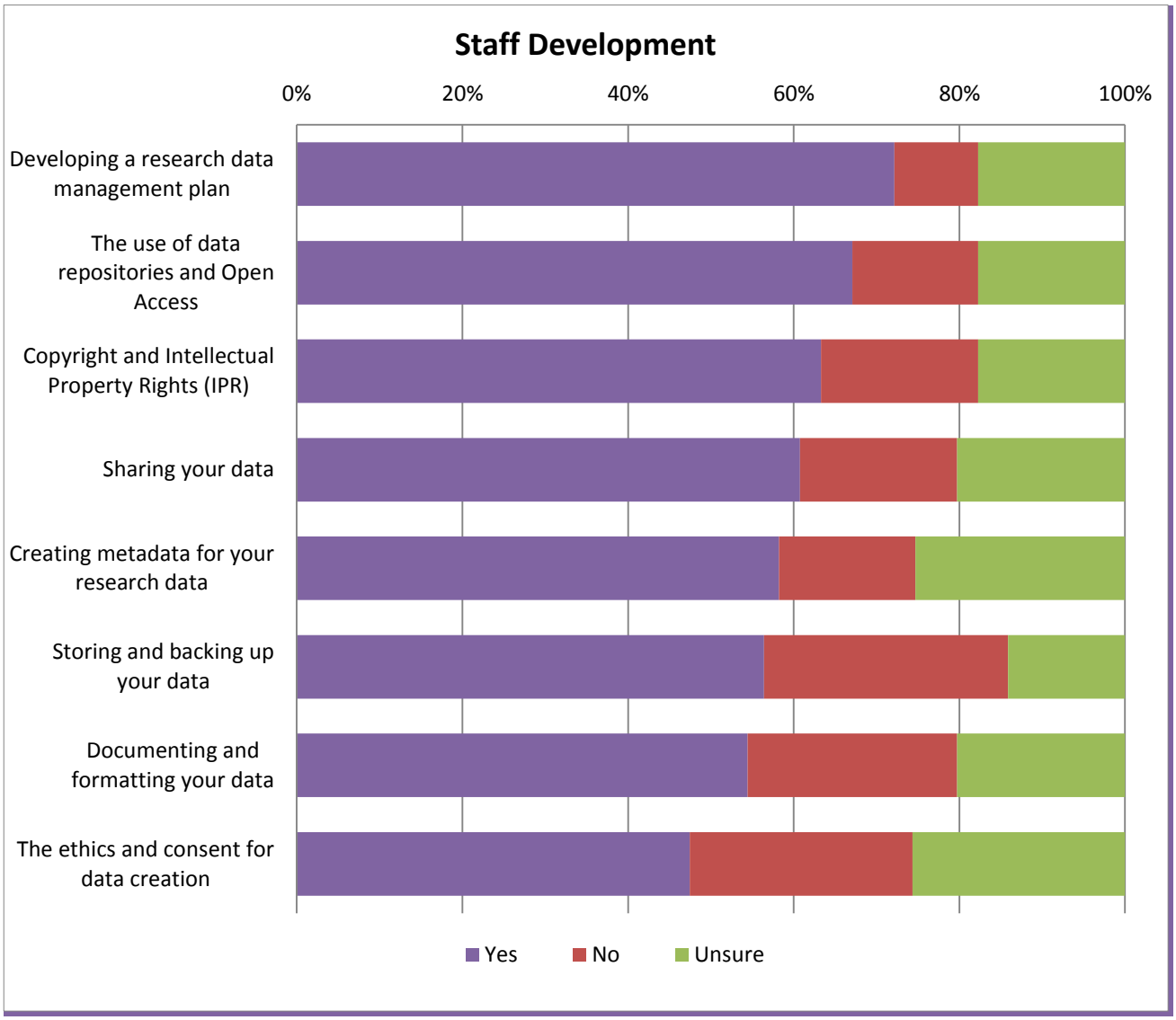
4.1 Have you received any training or support on research data management?

Only 10% of respondents have received any staff development on RDM. Of those who have received training, this came either from previous employment, as part of an MSc programme or self-study.

Answer Options	Response Percent	Response Count
Yes	10.1%	8
No	89.9%	71

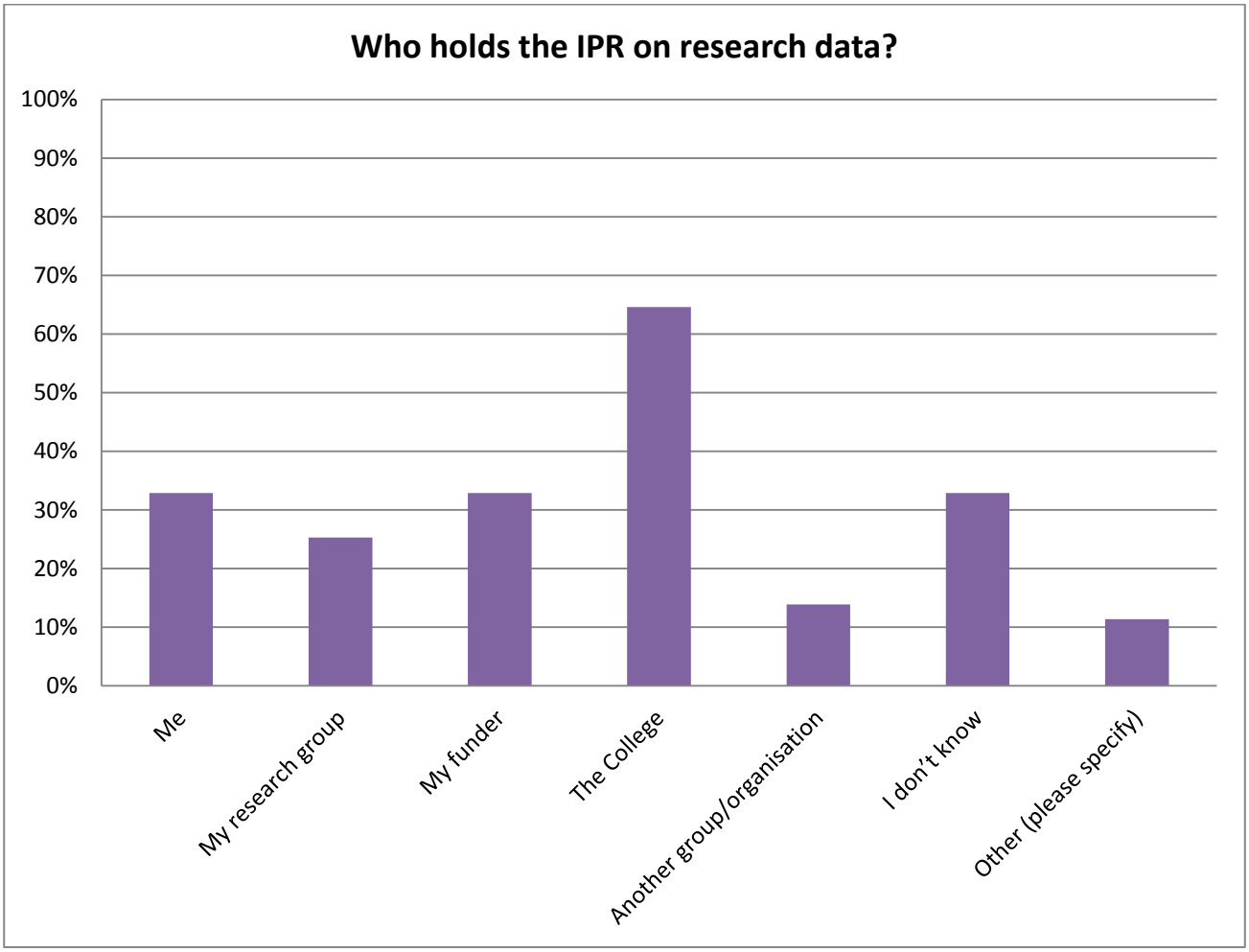
4.2 Would staff development or training be useful to you in the following areas of research data management?

Generally staff development was seen as useful on all subjects listed. RDM planning was seen as the most important (72%) with the ethics and consent of data collection seen as of lesser usefulness (47%).



4.3 Who has the Intellectual Property Right for your research data? Please select all that are appropriate.

When asked to identify who holds the IPR for the data, most users selected more than one option, but the most common answer was the College with 64%.



Appendix A – Copy of RDM Survey

Research Data Management Survey

This survey aims to gain a better understanding of the type of research data you produce and the areas of support required to help you to manage your research data. Responding to the survey will help us to us to plan and prioritise the support that should be provided to researchers within the College. The survey consists of 22 questions and should take you approximately 10-15 minutes to complete.

What is Research Data?

For the purpose of this survey, Research Data refers to any type of data created, collected, or generated that is analysed to produce original research results.

Who should complete this survey?

The survey is open to anyone who is actively involved in funded or unfunded research. This may include Principal Investigators, PhD researchers, research assistants and data managers.

General Information

1. Name
2. Please state which type of researcher you are:
    - Research fellow
    - PhD researcher
    - Research assistant
    - Post-doctoral researcher
    - Lecturer/researcher
    - Other (please specify)
3. If you are a PhD student, post-doctoral researcher or research assistant, who is your line manager?
4. Please list your department:
    - Comparative Biomedical Sciences
    - Clinical Sciences and Services
    - Pathology and Pathogen Biology
    - Production and Population Health
    - Other (please specify)
5. Who funds the research project(s) you are currently undertaking? If you receive multiple grants to fund your research list the top three funders.

Your research data

6. Please list the three most common tools you use to create and/or manipulate your research data (e.g. Stata. MySQL, MathWorks, Microsoft Excel).

7. What types of research data are created or worked with during the life cycle of the project? Please select all that are appropriate.
   - Electronic Documents
   - Spreadsheets
   - Databases
   - Questionnaires
   - Websites
   - Physical/Paper Notebooks
   - Audio/Video Content
   - Images/Photos
   - Slides/Specimens
   - Digital Objects
   - Raw Data
   - Models/Algorithms
   - Contents of an application
   - Other (please specify)

8. Where do you store your data during the life cycle of your project(s)? Please select all that are appropriate.
   - Your College supplied personal H:drive
   - A College shared drive e.g. R:drive or O:drive
   - Hard disk/local drive on College computer or laptop
   - Hard disk/local drive on personal computer or laptop
   - Instrument which generates data
   - Portable storage device e.g. external hard drive, USB stick
   - Project or user specific server e.g. NAS Box, PC shared drive
   - Via a service provided by a collaborating institution
   - CD/DVD
   - Email system
   - Web based service e.g. Dropbox, Google Drive
   - VHS/Video Cassette
   - Floppy Disk/ZIP Drive
   - Physical Photographs
   - Slides
   - Microfiche
   - On Paper
   - Other (please specify)

9. How frequently do you back up data created during the life cycle of your project(s)?
   - Daily
   - Weekly
   - Ad-hoc

- Never
- Don't know

10. If you do back up your data, where do you back it up to? Please select all that are appropriate.
    - College shared drive
    - External hard drive
    - USB/memory stick
    - CDs/DVDs
    - Self-managed server/NAS
    - With a collaborative partner
    - Cloud service
    - Don't know
    - Other (please specify)

11. Approximately how much data do you expect your project(s) to create during its life cycle?
    - Less than 1Gb
    - 1-50Gb
    - 50-100Gb
    - 100-500Gbs
    - 500Gbs -1Tb
    - 1-10Tb
    - 10-50Tb
    - Over 50Tb
    - I don’t know

    Additional comments

12. Are you aware of the backup process the College undertakes for data stored on its servers?
    - Yes
    - No

Research data management

13. Have you developed a Research Data Management Plan for any current research projects?
    - Yes
    - No
    - I don’t know

14. Are you aware of any specific research data management planning required from your funders?
    - Yes
    - No
    - I don’t know

15. Do you currently apply security checks on some or all of your research data?

- Yes (Please detail in question below)
- No
- I don't know

16. If yes, please detail any security checks you place on your data. Please select all that are appropriate.
    - Encryption
    - Anonymisation
    - Password protection of files
    - Access logging
    - Physical security (e.g. locked room)
    - Other (please specify)

17. At the end of a project, do you make your research data publicly available via Open Access/data repositories?
    - Yes (as a requirement of funding)
    - Yes (not required by funding)
    - No
    - I don't know

    If yes, please detail where and how you make your data available.

18. Do you document or record any metadata about your data, to make it more meaningful to others?
    - Yes
    - No
    - I don't know

    If yes, please specify standards or guidelines used.

Staff Development

19. Have you received any training or support on research data management?
    - Yes
    - No

    If yes, please specify where/how you received the information and the topics covered

20. Would staff development or training be useful to you in the following areas of research data management? (Yes, No, Unsure)
    - Developing a research data management plan
    - Creating metadata for your research data
    - Documenting and formatting your data
    - Storing and backing up your data
    - The ethics and consent for data creation
    - Copyright and Intellectual Property Rights (IPR)

- Sharing your data
- The use of data repositories and Open Access

Other (please specify)

21. Who has the Intellectual Property Right for your research data? Please select all that are appropriate.
    - Me
    - My research group
    - My funder
    - The College
    - Another group/organisation
    - I don't know

    Other (please specify)

22. Please provide any additional or general comments on research data management and its support within the College.