



# Research Data Management Survey

---

Authors:	Thomas Parsons, Shirley Grimshaw, Laurian Williamson
Audience:	Internal and JISC MRD partners
Published:	06/02/2013

## Contents

1.	Introduction .....	2
2.	Methodology.....	3
2.1.	Background.....	3
2.2.	Data collection.....	3
3.	Results .....	5
3.1.	Participants.....	5
3.2.	Types of research data created .....	8
3.3.	Data Storage.....	10
3.4.	The volume of research data .....	13
3.5.	Backing-up research data .....	14
3.6.	Using metadata to describe research data .....	18
3.7.	Externally funded research.....	18
3.8.	The development of a research data management plan .....	20
3.9.	Research data management training .....	21
3.10.	Accessibility of research data.....	23
3.11.	Intellectual Property Rights (IPR) for the research data .....	25
3.12.	Sensitivity of data and suitability for sharing .....	27
3.13.	Making research data publicly available .....	29
3.14.	Depositing data in a public subject/disciplinary repository .....	31
3.15.	Areas where help is required .....	33
4.	Conclusions.....	35
5.	Appendix .....	37
5.1.	RDM Survey questions and accompanying text .....	37

# 1. Introduction

The ADMIRE project is a JISC funded project<sup>1</sup> designed to create a sustainable Research Data Management infrastructure at The University of Nottingham (the University). The overall aim is to:

*"Establish and pilot a sustainable research data management (RDM) infrastructure for the University of Nottingham. It aims to develop an infrastructure to support the research data lifecycle, acknowledging & responding to differing practices across subject disciplines."*

As part of the requirements gathering phases, a survey was designed and disseminated to researchers across the University. This served multiple purposes:

1. To baseline current RDM practices
2. To gather the researcher's requirements for RDM
3. Raise awareness for the prospective service and gauge interest levels for the proposed service.

The survey covered typical aspects of RDM and provides a benchmark to measure progress against the Research Council UK's expectations for RDM. For example, the EPSRC<sup>2</sup> expectations mandate all funded research institutions to implement a support and technical infrastructure. This should enable:

1. Research data management throughout the research lifecycle
2. The publication and sharing of research data

This is the first survey of its kind at the University and will enable a clear view of current practice to be established, outline any gaps and identify areas for improvement. This will not only help to meet funding requirements, but will benefit researchers by the improvement of their day-to-day management of research data.

Furthermore, the University has always played a key role in promoting free access to a variety of materials; this includes access to teaching resources through 'Open Nottingham'<sup>3</sup> and a dedicated budget to fund Open Access publications. It is expected that the results of ADMIRE, will continue this tradition of openness and bring benefit to the wider community through the sharing of research data.

---

<sup>1</sup> <http://admire.jiscinvolve.org/wp/>

<sup>2</sup> <http://www.epsrc.ac.uk/about/standards/researchdata/Pages/expectations.aspx>

<sup>3</sup> <http://www.nottingham.ac.uk/open/opennottingham.aspx>

## 2. Methodology

### 2.1. Background

The survey questions were based upon the Digital Asset Framework (DAF) methodology<sup>4</sup>. Similar surveys have been carried out by Exeter<sup>5</sup>, Edinburgh and Northampton<sup>6</sup> using the DAF methodology, thereby allowing comparative analysis across institutions if required.

The survey instrument consisted of twenty questions in total. These included questions seeking demographic information, with a number of questions gathering richer data depending upon prior answers. Questions were multiple-choice (one answer), multiple-choice (multiple answers) and free comment. The analysis carried out upon completion was both qualitative and quantitative.

A copy of the survey is included in the Appendices.

### 2.2. Data collection

The collection of data was via an online survey using the Bristol Online Survey (BOS) tool<sup>7</sup>. The University holds a subscription to the BOS tool that allows the University logo to be used and a URL that includes Nottingham<sup>8</sup>; thereby reassuring participants that this survey was for University researchers and not a public survey.

A small pilot group of ADMIRe project members and University researchers acted as testers for the survey design. Changes were made to the questions based upon their feedback and this ensured accuracy before sending to a wider audience. The wider survey sample consisted of:

1. Career researchers (i.e. Lecturers, Research Fellows, Professors)
2. Post graduate researchers (not taught courses)

As there is no central means to distribute surveys, a combination of tactics were used to target researchers with University of Nottingham email accounts. A targeted email and corresponding HTML email template<sup>9</sup> were devised, so as to emphasise the professionalism and importance of the survey. An incentive in the form of an iPod Shuffle was also used to encourage responses, with one person chosen at random to win the prize.

---

<sup>4</sup> <http://www.dcc.ac.uk/resources/repository-audit-and-assessment/data-asset-framework>

<sup>5</sup> [https://eric.exeter.ac.uk/repository/bitstream/handle/10036/3689/daf\\_report\\_public.pdf?sequence=1](https://eric.exeter.ac.uk/repository/bitstream/handle/10036/3689/daf_report_public.pdf?sequence=1)

<sup>6</sup> <http://nectar.northampton.ac.uk/2736/>

<sup>7</sup> <http://www.survey.bris.ac.uk/>

<sup>8</sup> <http://www.survey.bris.ac.uk/nottingham/data/>

<sup>9</sup> <http://www.nottingham.ac.uk/emailTemplates/researchdatamanagement.htm>

The distribution of the survey was via four principle methods in a two stage phased approach:

#### Phase I (5<sup>th</sup> July to 14<sup>th</sup> August 2012)

1. Sent via Heads of Schools, School Managers and RDM 'champions' who were already engaged with the project
2. Sent to post-graduates via the Graduate School and research support officers
3. Distribution via Faculty Librarians and IT Support staff
4. Distribution across the University via the Message of the Day (a pop-up that displays when users log onto the network)
5. Distribution via departmental newsletters and website news sections

#### Phase II (16<sup>th</sup> August to 14<sup>th</sup> September 2012)

1. Via an email distribution list to all career researchers (obtained via Human Resources)

In total the survey ran from 5<sup>th</sup> July 2012 to 14<sup>th</sup> September 2012. The original planning estimated the survey would close after one month, but increased awareness of the project and enthusiasm from senior management, allowed the phased approach to be adopted and a targeted email shot to be used. As with any survey, choosing a launch date is problematic - although the summer months are traditionally a holiday period, it was decided that refraining from sending the survey until later in the year would clash with the start of the academic term. So instead we chose to extend the survey close date, therefore allowing those on holiday sufficient time to respond should they wish.

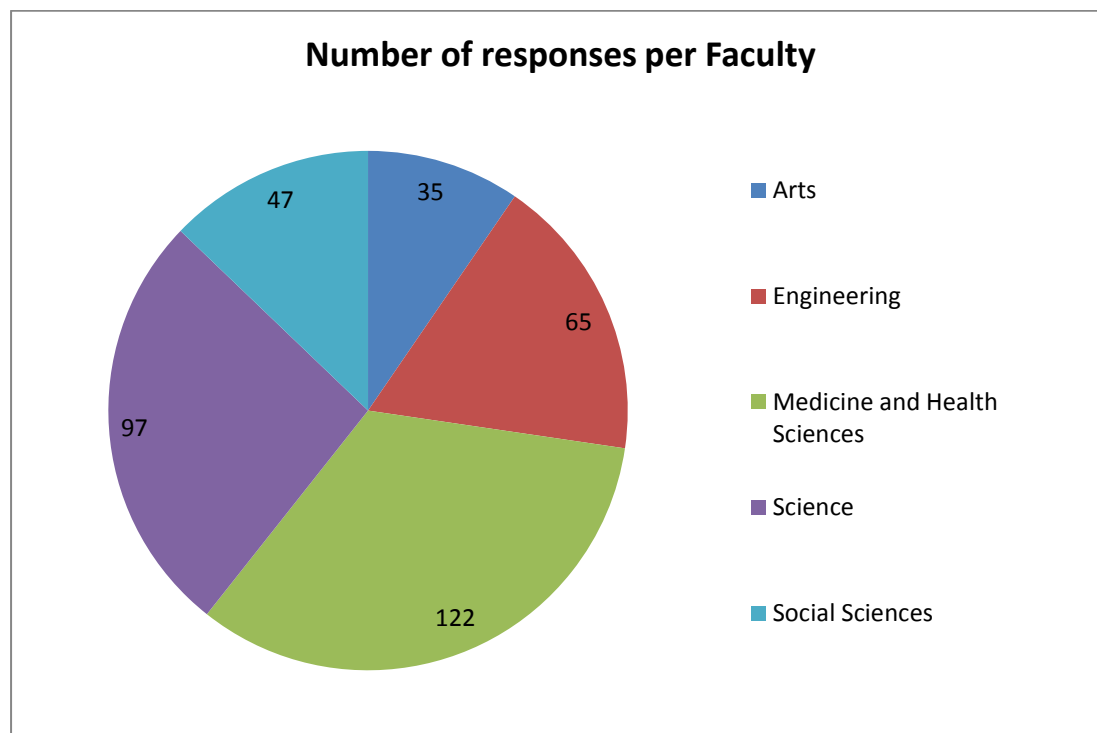
In total 366 researchers responded, a figure which was highly encouraging for the project team and allowed a detailed analysis to be made.

### 3. Results

The following section summarises the key findings of the study and helps to shed light on both current RDM practice and the requirements for the new RDM service.

#### 3.1. Participants

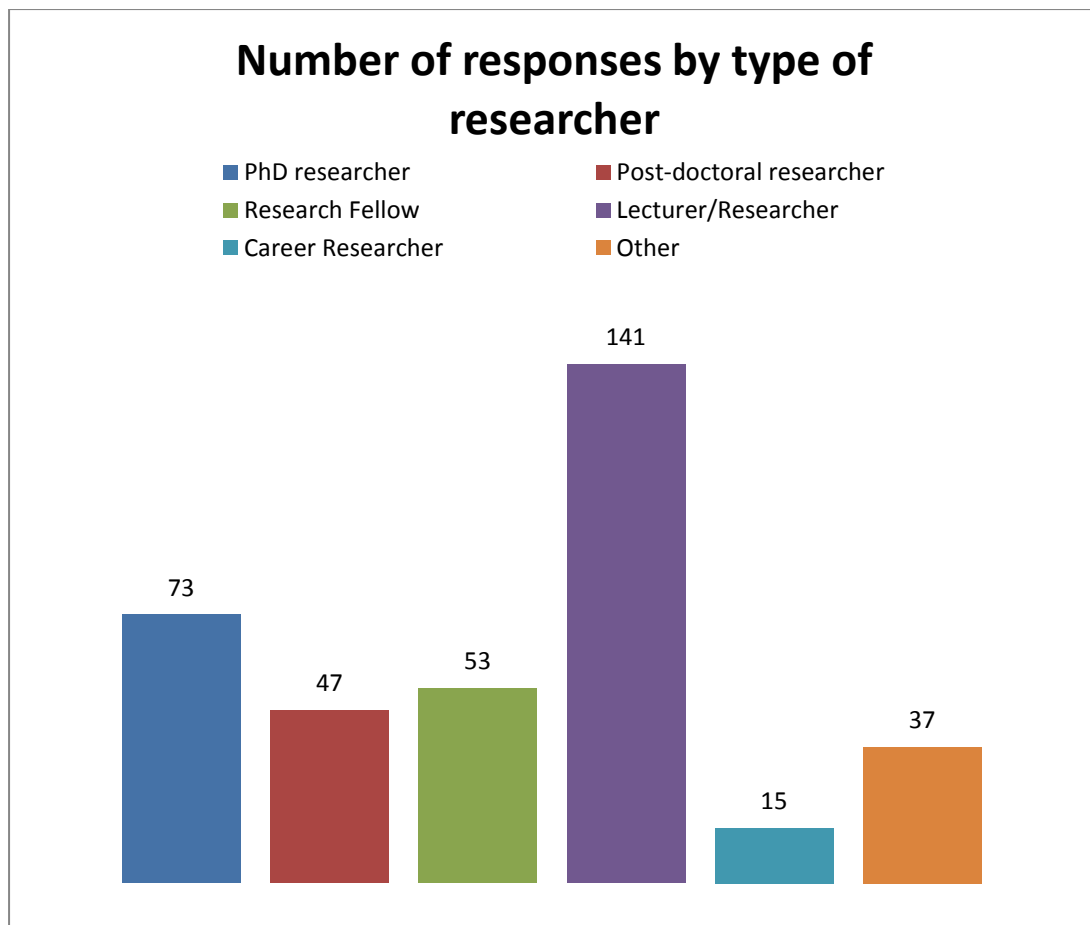
Altogether 366 people responded to the survey and the number of respondents from each of the Faculties is shown below:



**Figure 1: Number of responses per Faculty**

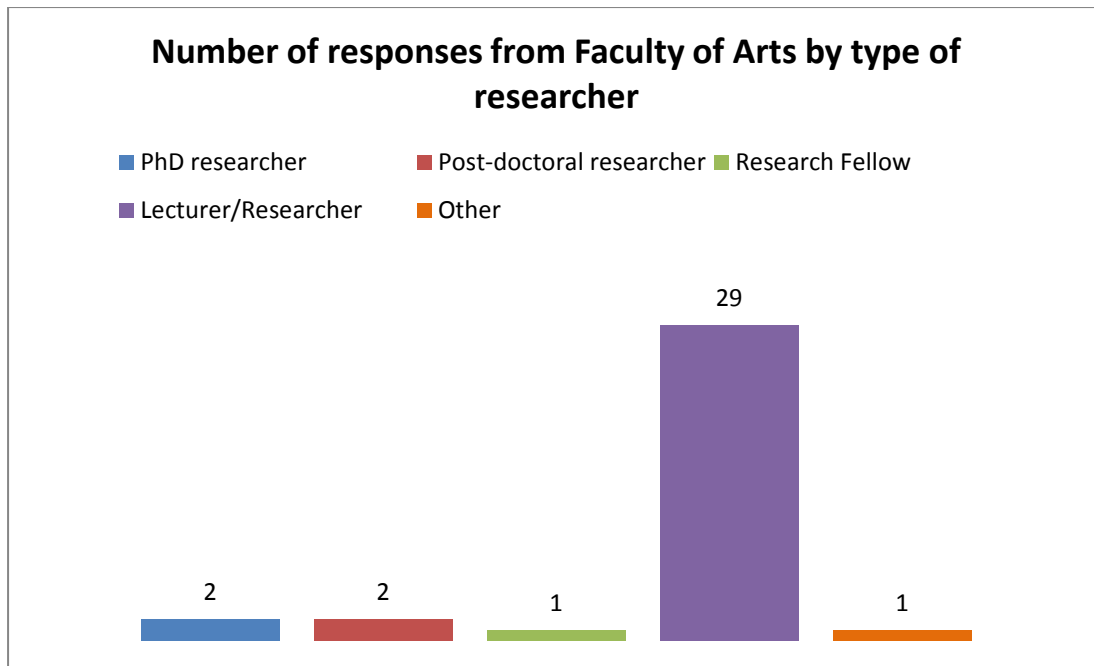
As can be seen from Figure 1, the largest group of respondents were from Medicine and Health Sciences, followed by Science and Engineering. The Faculty of Social Sciences and, in particular, the Faculty of Arts had fewer respondents. It is worth clarifying that these results are representative of the sizes of Faculties, with Medicine and Health Sciences (MHS) and Science being by far the largest research Faculties within the University.

Looking at the type of researcher who have responded to the survey, it is clear from Figure 2, that most were either Lecturers/Researchers or PhD researchers:

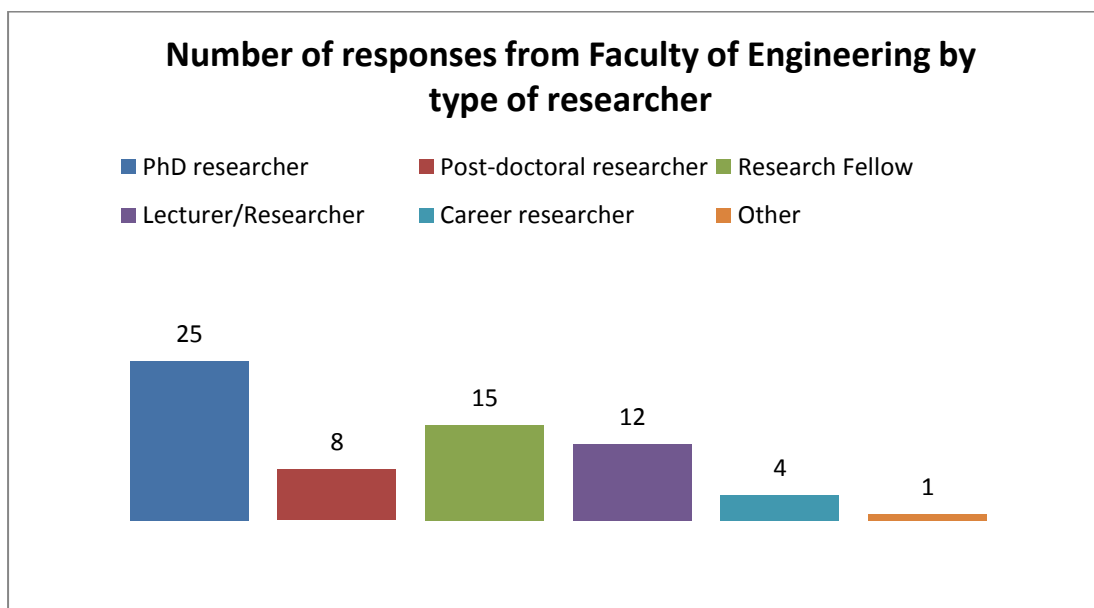


**Figure 2: Number of responses by type of researcher**

In all Faculties, apart from Engineering, the largest group of respondents is Lecturers/Researchers. In Engineering the largest group is PhD researchers. However, when this is analysed more closely, it can be seen from Figure 3 – Figure 4 that the type of respondent does vary across Faculty. This may indicate that the sense of involvement with RDM differs across the Faculties and with respondent types; or that simply the channels used to distribute the survey were stronger in certain areas over others.



**Figure 3: Number of responses from Faculty of Arts by type of researcher**



**Figure 4: Number of responses from Faculty of Engineering by type of researcher**

The number of respondents who selected 'Other' to describe their research role also varied greatly across Faculties. The most common roles in the 'Other' category were given as Professor, Assistant Professor, Research Assistant, Research Associate, Research Officer, Technician and Manager of some kind.

It appeared that some respondents preferred to describe their role in terms of their official title within their School or Department. As the range of titles and roles may be larger in some Faculties than in others this may explain the greater use of the 'Other' category by those Faculties.

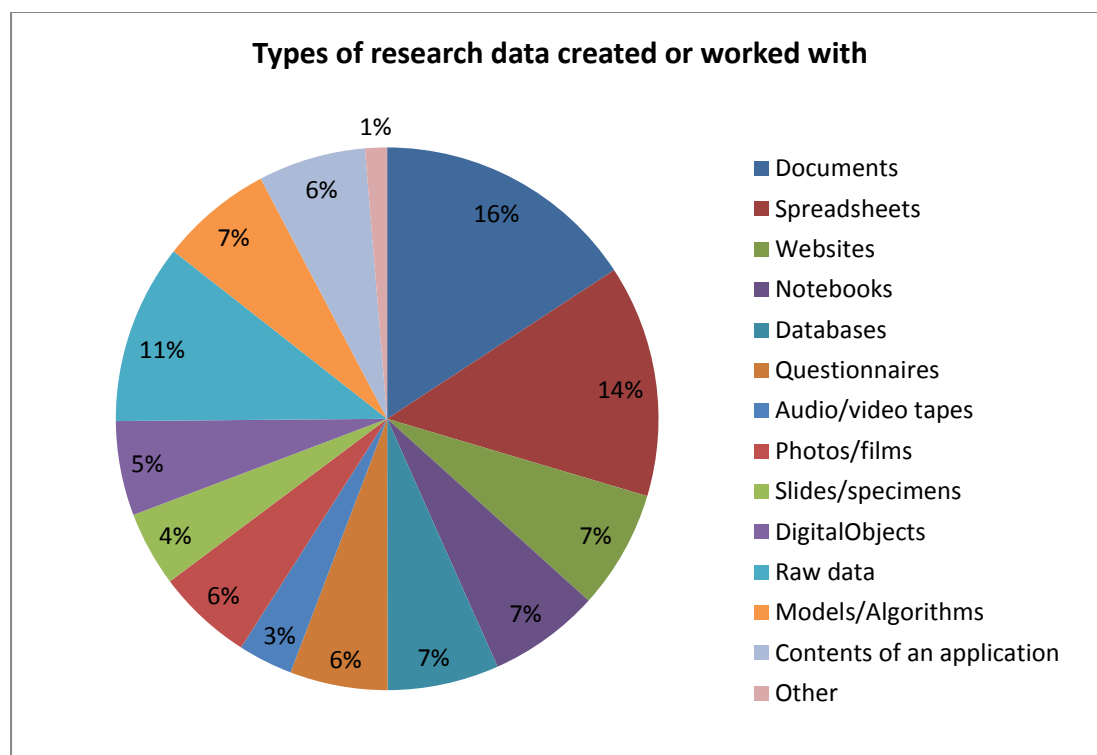
In order to understand the demographics in more detail, question 5 was an optional question where respondents could summarise their main research work. Thereby providing valuable contextual information about:

- Differences and commonalities in practice between research groups working on similar areas
- The ability to highlight projects that fall under the University key research themes<sup>10</sup>

The last point is important at a project sustainability and RDM business case level. Supporting and highlighting the need for RDM services for high profile projects, aligns the project with the University strategies and therefore, can be used to add weight to the argument for providing such a service. Clearly, this does not diminish the importance of providing RDM support to all projects, but it does help at a political level when making the case for a new RDM service.

### 3.2. Types of research data created

Question 6 asked respondents to identify the types of research data they created as part of their research. Figure 5 shows that documents, spread sheets and raw data were identified by the highest percentage of respondents:



**Figure 5: Types of research data created or worked with**

<sup>10</sup> <http://www.nottingham.ac.uk/research/priorities/index.aspx>



Only 1% of respondents identified 'other' data, thereby confirming that the majority of data types were represented and known to the project team. Responses in the other categories were: software scripts or code, images - both digital (MRI Dicom files) and physical (manuscripts or architectural drawings) and DNA sequence data.

The term 'raw data' includes "*Raw data files generated by software, sensors or instruments files*". Comments and follow-up interviews, suggests that many of these files are generated by laboratory machines running bespoke software on older operating systems, some of which will be networked and some not. Although these are predominantly found within the Physical Sciences, the Arts Faculty does generate this data as well, with one researcher citing 3D scans of historical statues as their primary research data.

While the majority of data is digital, significant amounts of data is held within physical notebooks or lab notebooks in particular. Follow-up interviews have highlighted that lab notebooks are widely used across the University to record experiment parameters, methodologies and results.

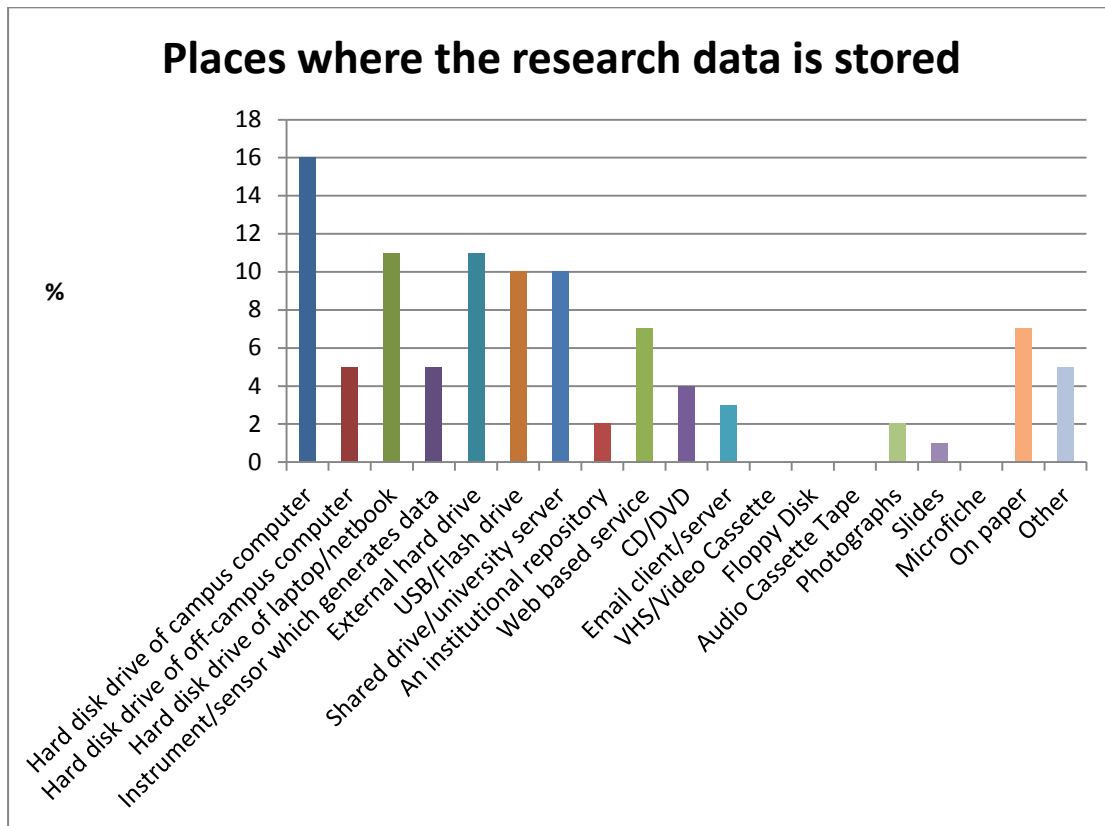
Physical data from other Faculties e.g. historic papers, questionnaires, newspapers, manuscripts or ethnographic data such as correspondence letters, suggests there is a need to manage or at least catalogue this type of data. Particularly when RCUK policies (EPSRC Point IV<sup>11</sup>) include reference to providing access to non-digital data; the assumption being that non-digital data may have to be digitised before sharing. Follow-up interviews with researchers confirm that many are scanning notebooks to PDF (in particular lab notebooks), yet the majority do not. Further complications would arise if the non-digital object held copyright connotations, licencing restrictions or sensitive data.

---

<sup>11</sup> <http://www.epsrc.ac.uk/about/standards/researchdata/Pages/expectations.aspx>

### 3.3. Data Storage

Question 7 sought to understand how this data is being typically stored. This was a multiple choice/multiple answer question and respondents were asked to tick all that applied. Figure 6 shows the percentage of respondents who stored their research data in the different places suggested:



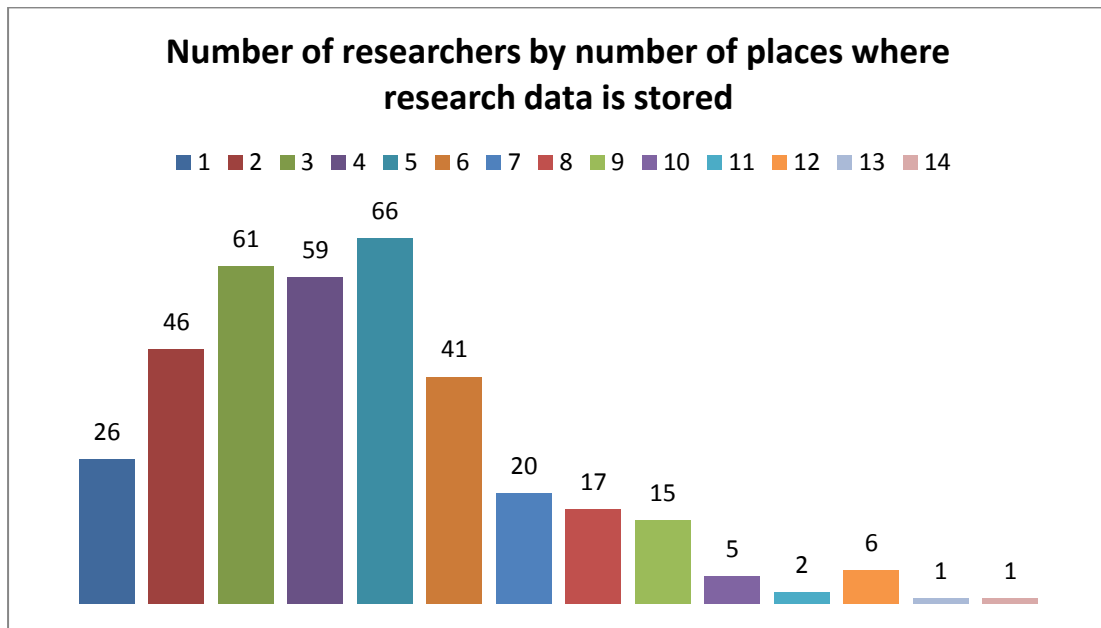
**Figure 6: Places where the research data is stored**

The percentages offer insight into typical behaviour towards storing data. Strongly indicating that data is stored in multiple locations, with campus computers, laptops, external hard drives, USB drives, University storage, web based storage or paper being the top answers. The "other" category was used by many to expand upon these answers, with "DropBox" being the typical response and many mentioning the University's Attix backup software being used.

Other places mentioned were Google Drive, Microsoft Skydrive, YouTube, a departmental server and a University managed Microsoft Sharepoint server. Paper copies as well as audio copies were also mentioned.

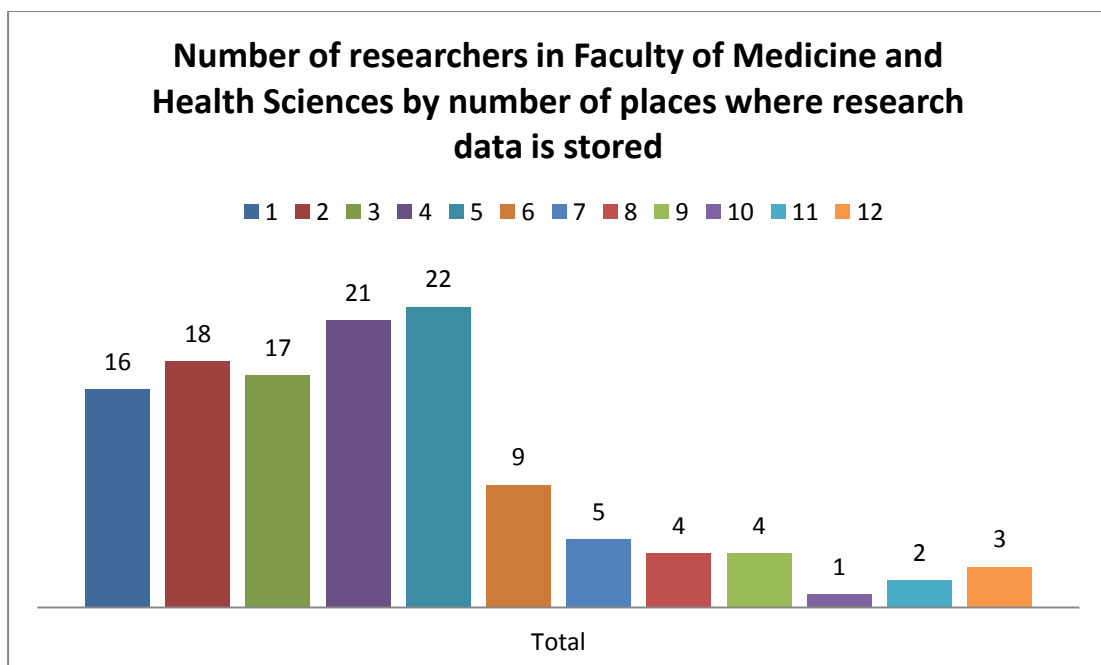
While this analysis provides an overall view of storage behaviour, it does not tell us whether respondents saved their data to more than one place, but judging by the array of responses this is suggested. In order to establish this, the data was analysed a little deeper.

Figure 7 shows the number of places in which respondents store their research data:



**Figure 7: Number of researchers by number of places where research data is stored**

The majority of Faculties fit the pattern shown in Figure 7. However, Figure 8 shows an interesting difference for the Medical & Health Sciences Faculty:



**Figure 8: Number of researchers in Faculty of Medicine and Health Sciences by number of places where research data is stored**

It can be seen that a greater proportion of respondents from Medicine and Health Sciences stored their data in just one place. To establish how great a concern this might be, the survey data was examined further to establish where in particular data was stored when only one place was chosen. Storing data to just one place is obviously a far lesser risk when that place is a university server where data is backed up on a regular basis, the converse would be storing data to removable media such as a USB drive. The results by Faculty are shown below in Table 1.

**Table 1: Sole places where respondents from different Faculties stored their research data**

	Arts	Engineering	Medicine and Health Sciences	Science	Social Sciences
Laptop	1				
Hard disk drive of campus computer		2	3	1	1
External hard drive		1			
Shared drive/ university server			10	2	1
Hard disk drive of off-campus computer			1		
USB Flash drive			1		
Floppy Disk			1		
Hard disk drive of laptop/netbook					1

The results are encouraging, in that they illustrate that a far higher proportion of Medicine & Health Sciences Faculty (MHS) researchers are storing their data on secure and backed-up media (i.e. University file storage) and therefore automatically complying with regulations governing the storage of sensitive data. This requires further analysis to confirm, but for many, it may be one of the few places where it is ethically and legally acceptable to store their data. Indeed, it is likely that in order to get a research project approved in the first place, many MHS researchers will have specified where their data will be stored.

For examples, the Medical Research Council has had such a policy in place since 2005<sup>12</sup> and requires that data is actively curated throughout the research lifecycle. Follow-up interviews with MHS researchers confirmed that the issue of Data Protection and patient confidentiality forbids them from storing data in many typical places seen in other Faculties, for example using DropBox or USB memory sticks.

Although MHS may indicate a preference for University storage, a further analysis of the overall results showed that only 173 respondents used the University networked storage, which equates to 48% in total. This is surprisingly low; given that all researchers have a networked drive allocated by default to their account and will be aware of this through their staff induction processes. However, this currently only provides 4 GB of storage, so the following questions regarding volumes of data may shed more light upon this. Larger storage is available through shared network drives, although access to folders is usually determined at a group level and not to the individual.

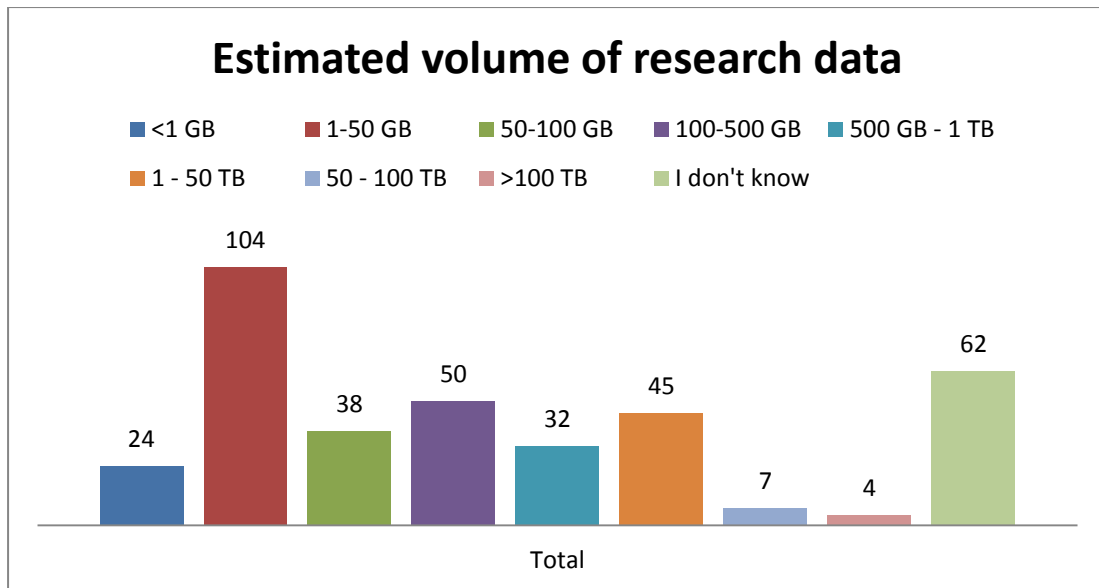
It should be noted that mid-way through the survey, University IT Services deployed a "Research Filestore" service which offers 1TB of: free, secure, backed-up and networked storage for any researcher across the University. This facility was not mentioned by any respondents though, so a repeat of this survey in one year may see the reliance upon University provisioned storage increase.

### 3.4. The volume of research data

Question 8 asked respondents to estimate the volume of research data they created across all of their work. Figure 9 shows the number of respondents who placed the volume of their research data into the particular categories provided.

---

<sup>12</sup> <http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/datasharing/Policy/index.htm>



**Figure 9: Estimated volume of research data**

By far the greatest number of respondents estimated the volume of their research data to be between 1 and 50 GB. Only a small number of respondents estimated the volume of their research data to be greater than 50 TB and a considerable number had no idea of the volume of data they were creating. The Faculties of Arts and Social Sciences had lower estimates of volume of research data than the Faculties of Engineering, Medical and Health Sciences and Science. In general the breakdowns were remarkably similar, excepting a spike in the number of researchers in the 1-50TB category for the Sciences.

The results can be used to identify where greater data storage services are required, but they do not indicate why the usage of the networked file services is not near 100%. A tentative conclusion would be that a typical researcher requires 1-500 GB, with some users requiring significantly more in the Sciences, Engineering and MHS Faculties.

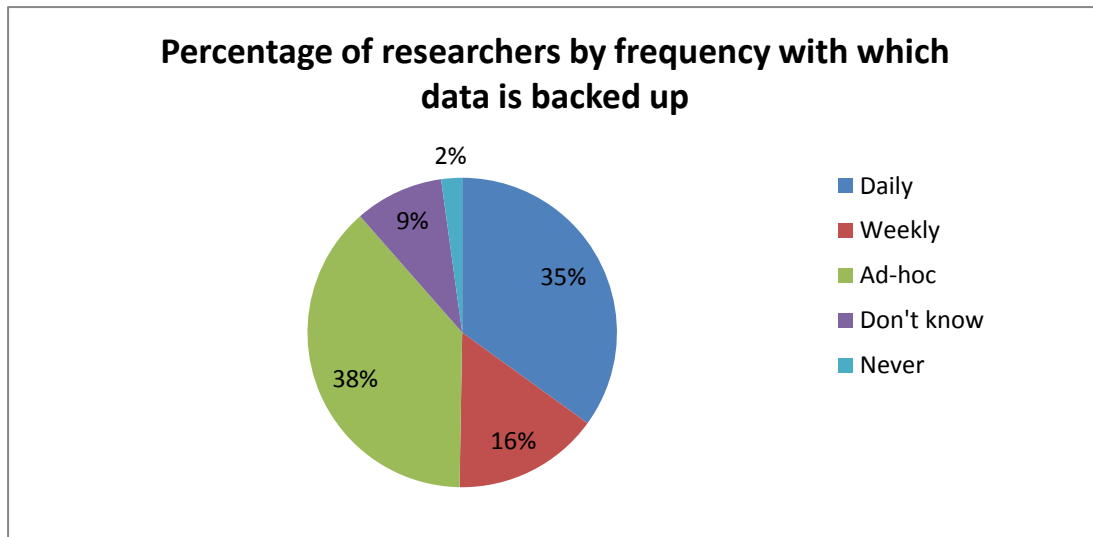
Additional follow-up focus groups categorically state that researchers are against deleting any of their data, so more work will be required to understand the capacity for storage, archive and preservation of data in the future in line with the 10 year retention period specified by RCUK.

### 3.5. Backing-up research data

Respondents were asked how frequently they backed-up their research data. Figure 10 shows the responses obtained.

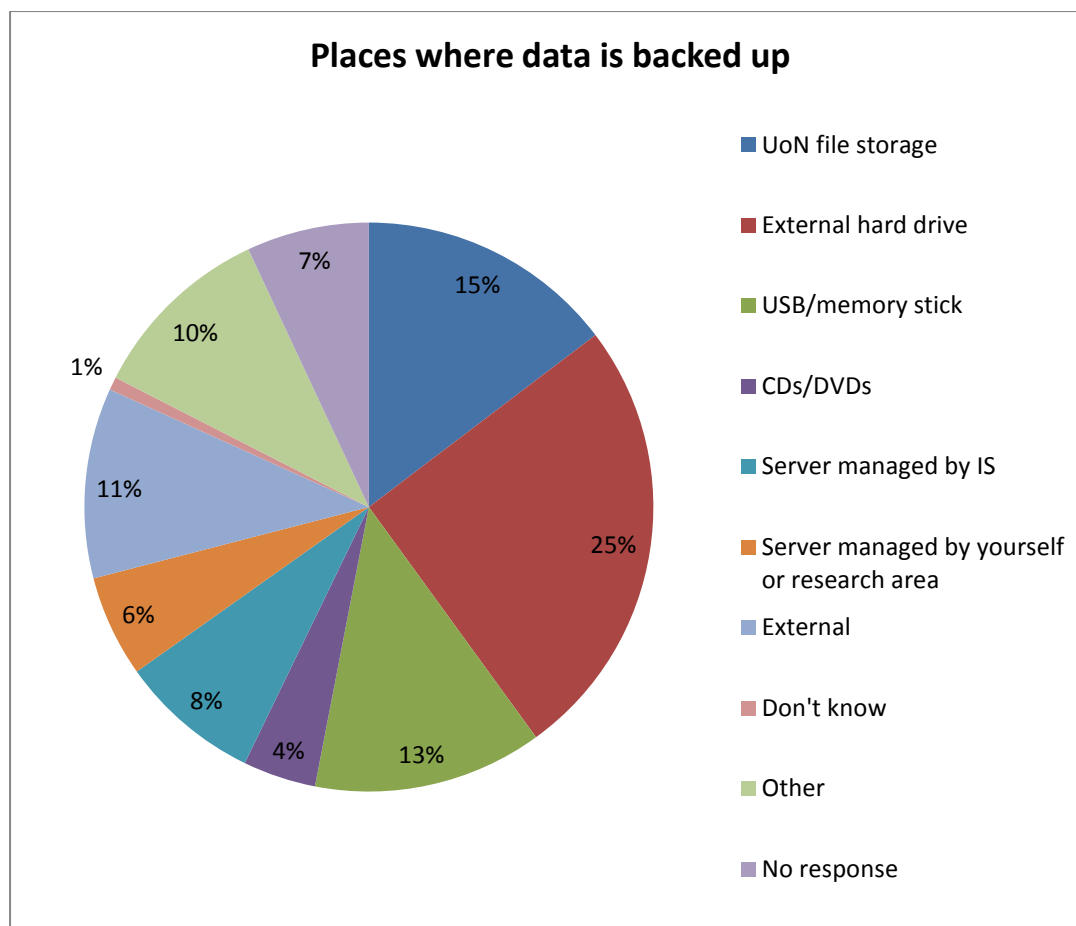
Only 35% of respondents backed up their data on a daily basis. For a large percentage of respondents, backing up of data was not done regularly, with another 9% admitting that they did not know when it was backed up and 2% admitting that they never backed it up at all.

While on the surface this appears worrying in terms of data management, it may well be that those who did not know when it was backed up and those who claimed never to back up their data answered thus because their data was backed up automatically on a shared server. To obtain a more accurate picture of this, further analysis would need to be carried out.



**Figure 10: Percentage of researchers by frequency with which data is backed up**

Respondents were also asked to state where the data was backed up. Figure 11 shows the percentage of respondents who backed up their data to specified places.



**Figure 11: Places where data is backed up**

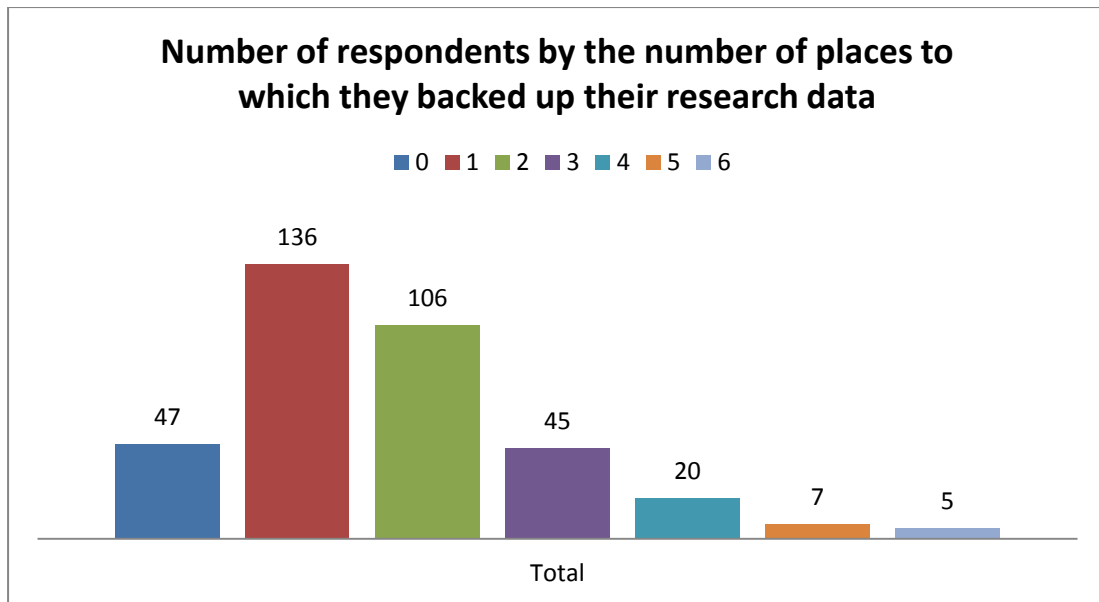
The majority of respondents backed their data to an external hard drive. The next most popular place for backing up data was UoN file storage.

A variety of answers was provided in the 'Other' section. Again the most frequent places quoted for backing up data were Dropbox and the University backup software named Attix. Other places mentioned were Apple's Time Machine (which presumably backs up to an external hard drive), email attachments, and printed copies. Many respondents chose to back up their data to multiple PCs, laptops, external hard drives and USB sticks. Remote servers and other cloud services were also mentioned.

As when analysing responses to data storage, a more accurate picture of 'backing up' behaviour can be obtained by looking at the number of places to which respondents back up their data. Figure 12 shows the number of places to which respondents backed up their data.

As can be seen from the Figure 12, most respondents backed up their research data to either one or two places. A few respondents backed up to as many as six places.



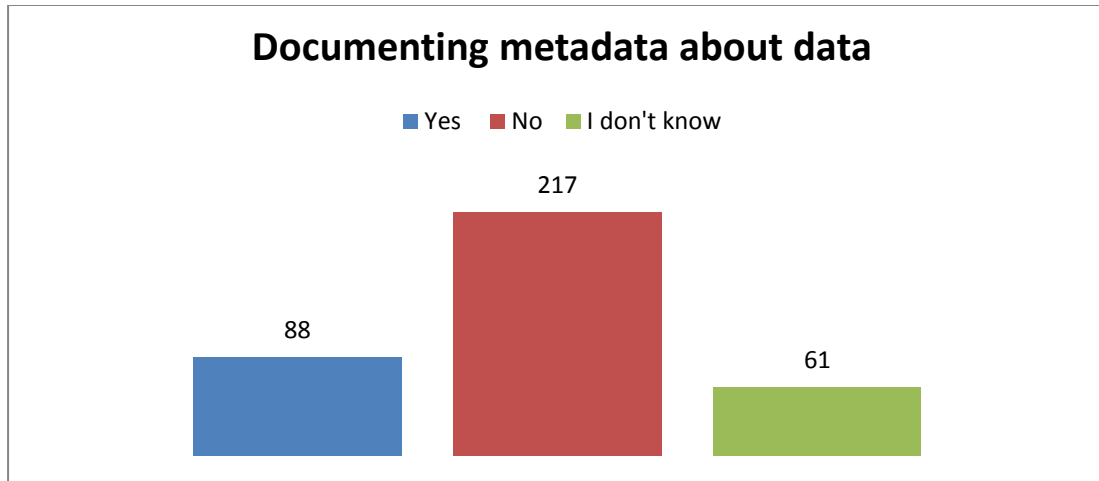


**Figure 12: Number of respondents by the number of places to which they backed up their research data**

The use of external hard drives is widespread and interview follow-ups explain that they are cheap and readily available. Unfortunately as one respondent noted, you really need to back-up the external hard drive as well, as they are prone to failure or loss, particularly if the researcher travels frequently. Although this survey did not explicitly ask the nature of the data that is backed-up, no respondents commented on whether they used versioning or deleted old copies of files of datasets. Further work is needed to understand whether back-ups are simply a copy of a directory “as-is” or more selective back-up of only files that have changed.

### 3.6. Using metadata to describe research data

Question 10 asked respondents whether or not they recorded metadata about their research data in order to make it more meaningful and easier to search for. Figure 13 shows the responses:



**Figure 13: Documenting of metadata about data**

The majority of respondents did not record metadata. Those that did were asked whether they used any standards or guidelines when constructing it. Of the 88 respondents who said they constructed metadata to describe their research data, only 16 used any standards or guidelines when constructing the metadata. All responses, apart from the first response, described metadata standards. The first response seems to be describing standards procedures for entering data, thereby illustrating a need for metadata training awareness across the University at a minimum.

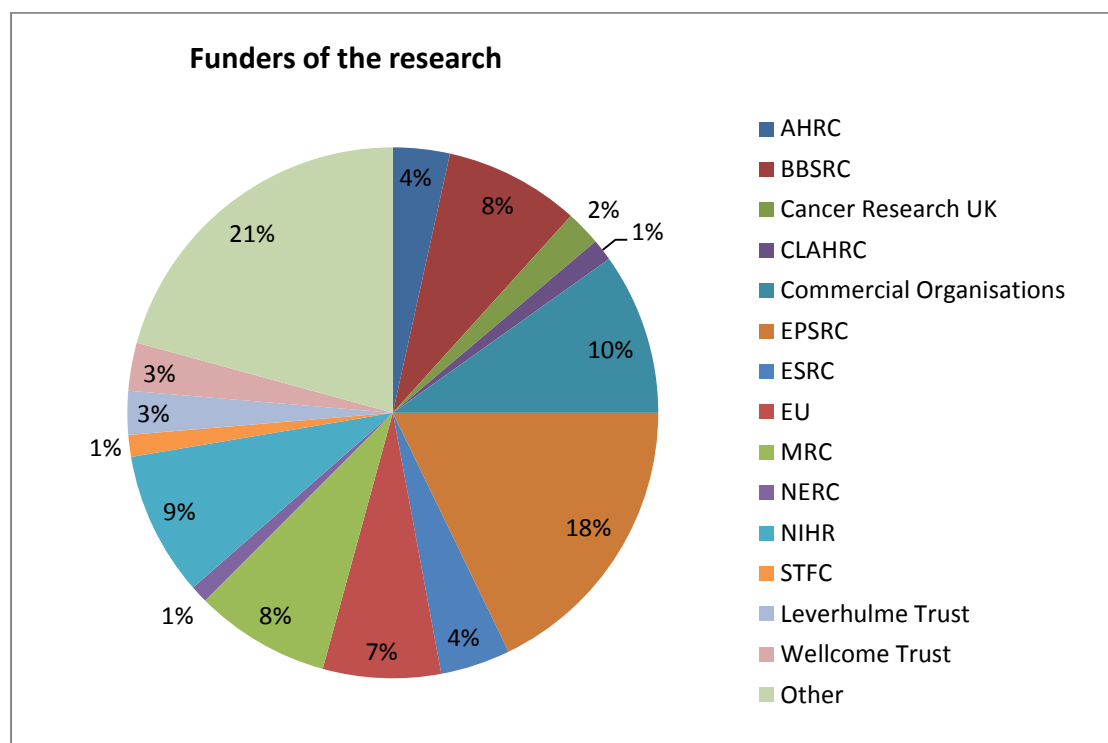
In order to meet the funding requirements mandates for cataloguing datasets, metadata must be used to describe data. These results suggest that this will require training across the University and a cultural shift for most people – it is safe to say, that creating metadata is simply not in the research workflow of the majority of researchers. In order to gauge the extent of how much training activity will be required, the following questions sought to understand who funds current research and what level of support should be expected.

### 3.7. Externally funded research

As mentioned previously, the shift in funders' requirements regarding data sharing now requires the majority of publicly funded research projects to create a Data Management Plan (DMP) and to share their data. Costing and resourcing a new RDM service at the University requires an understanding of how researchers are typically funded. High-level data is available concerning grant allocations, the number of researchers on the payroll, yet this doesn't fully cover the actual number of researchers generating data as part of this work - particularly if projects utilise associates, collaborators, students or temporary staff.

Therefore in order to estimate what level of support is likely to be required in these important areas, the survey asked respondents whether or not they were currently working on an externally funded research project. They were then asked to name their funders.

The survey results show that 71% of respondents are currently working on a research project that was externally funded, with 29% being internally funded.



**Figure 14: Funders of the research**

As can be seen from Figure 14, the range of external funders is quite large. Many will fall under the RCUK remit and have similar data management requirements, but many other such as Cancer Research UK<sup>13</sup> or other charities, will have their own RDM policies that require data sharing.

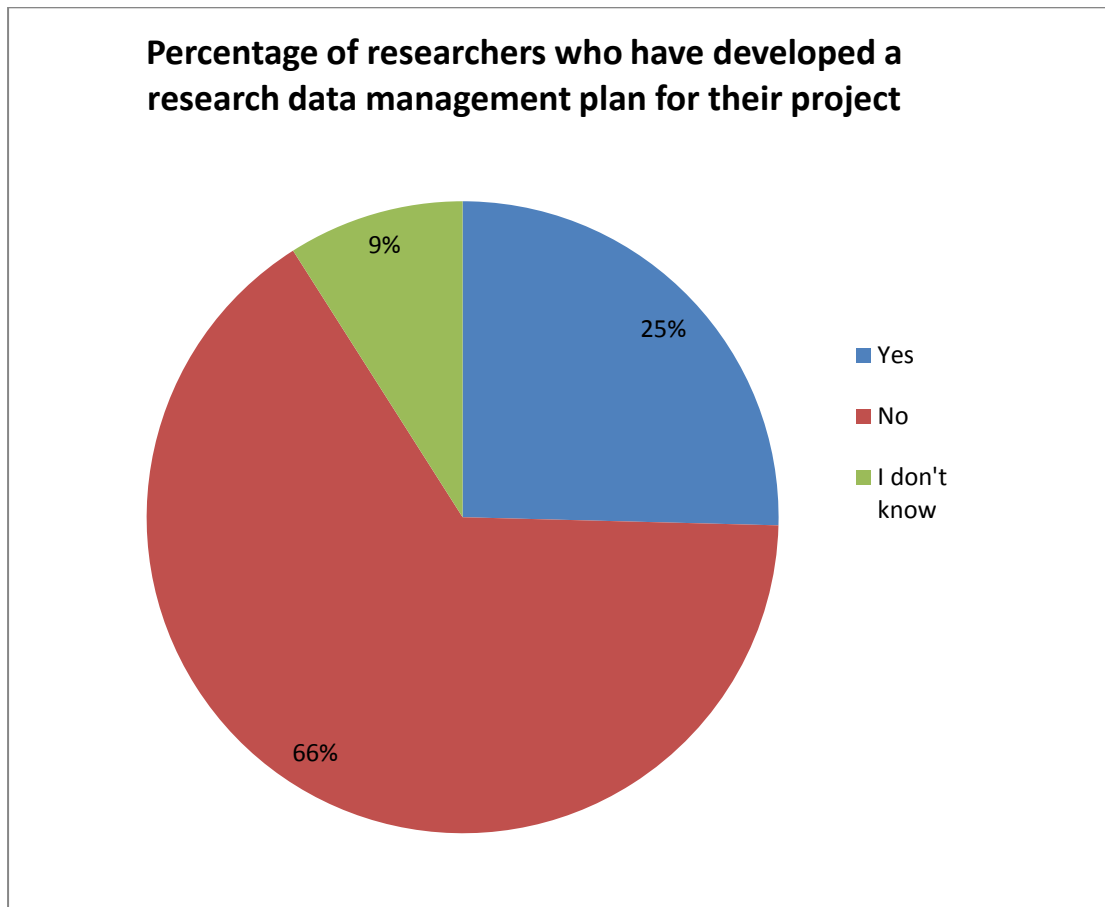
Fifty six different funders were mentioned by name, plus several unnamed charities and industries. The huge variety of funders will, of course, have implications for the training, guidance and advice required by researchers when it comes to: on-going data management, the grant application stage and data sharing.

<sup>13</sup> <http://www.cancerresearchuk.org/science/funding/terms-conditions/funding-policies/policy-data-sharing/>

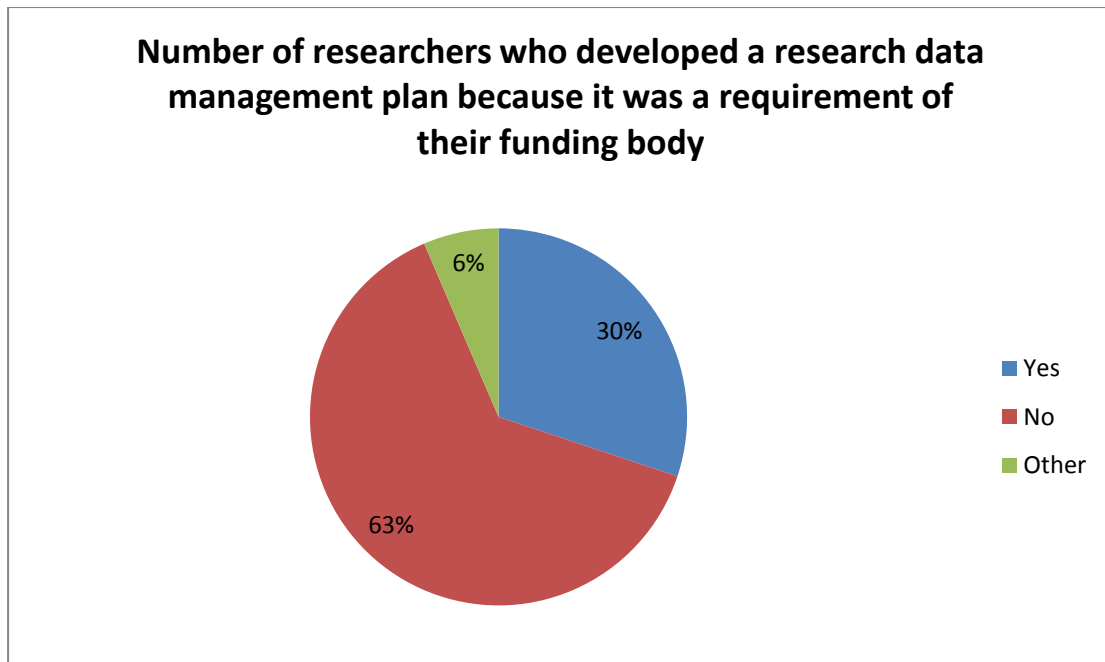
### 3.8. The development of a research data management plan

In order to estimate current practice around creating DMPs, respondents were asked in question 12 whether or not they had developed a research data management plan for their project, and if so, whether it was in response to data requirements set out by their funding body.

Figure 15 shows the percentage of respondents who had developed a plan. Figure 16 shows whether or not the development of a research data management plan was in response to the requirements of their funding body.



**Figure 15: Percentage of researchers who have developed a research data management plan for their project**



**Figure 16: Number of researchers who developed a research data management plan because it was a requirement of their funding body**

Further analysis of the data showed that Research Fellows, Career researchers and 'Other' researchers (Professors, Assistant Professors, Research Assistants, Research Associates, Research Officers, Technicians and Managers) were most likely to have developed a research data management plan than post-doctoral researchers or PhD researchers. As suspected, this indicates that those at Principal Investigator (PI) level are creating DMPs as they are the ones applying for the funding, yet those working on the project or under the PIs supervision are not. Further work is required to understand this and particularly to understand if the project members are aware of the DMP or not.

Figure 16 does illustrate that a surprising number (59 in total) developed DMPs independently of funding requirements, thereby suggesting that this may be standard practice amongst some researchers within the University. Two comments confirmed that this is standard practice when conducting Clinical Trials or when gaining ethics approval.

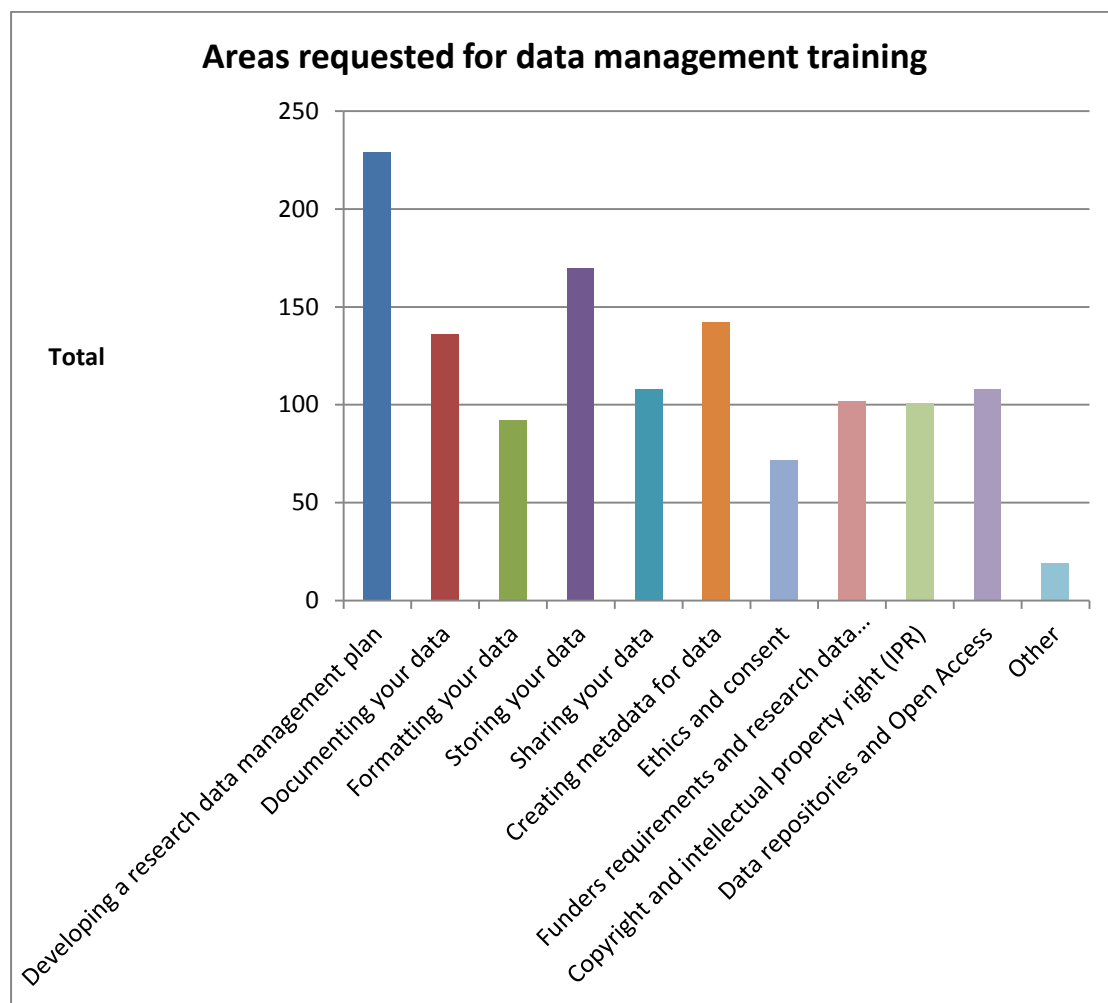
### 3.9. Research data management training

Question 14 found that the percentage of respondents who had received research data management training was only 7% of the total sample. When asked to name the training they had received in research data management there seemed to be some confusion about what was meant by the question. Some respondents answered it in terms of when they had received training – usually as part of an undergraduate or postgraduate degree; others gave the name of the training body or course. Those who did describe the training they had received listed the following:

- Good clinical practice training

- Stats training
- Data collection techniques
- Data organisation and management
- Documenting data
- E-workbook training
- Research methods with some data management elements
- Research data storage
- Ethics/regulatory approval/Data Protection Act
- Intellectual Property Right
- Back up

Two respondents answered that they had acquired their knowledge over the years (20+ years) as part of their jobs. The results and comments indicate great opportunity for training to be provided, and also highlight the ambiguity over what could constitute RDM training. Respondents were then asked to tick from a selection of 10 areas of data management training any that they would like to receive. Responses are shown in Figure 17.



**Figure 17: Areas requested for data management training**

The greatest demand was for “Developing a research data management plan”, followed by “Storing data”, “Creating metadata for data” and “Documenting data” (details of methodology, equipment used, details of physical specimens etc.). How much these priorities had been influenced by the previous questions on the survey is hard to determine. There was little request for Ethics and consent – possibly because this is already dealt with by other established courses and training provided by the University and Schools.

Only 4% of respondents said that they wanted no training, which is positive in terms of service acceptance and interest. One respondent commented *“this is critical information for any researcher, it should form part of an induction package for new research staff at any level.”*

Comments added to the ‘Other’ section requested training on:

- Handling data protection and freedom of information requests
- Exploring higher levels of manipulation of data
- How to extract metadata from data file formats
- The various storage options that are available and importantly, who is responsible for this
- IT training directed specifically towards the requirements of an Analytical Facility supporting a large School.

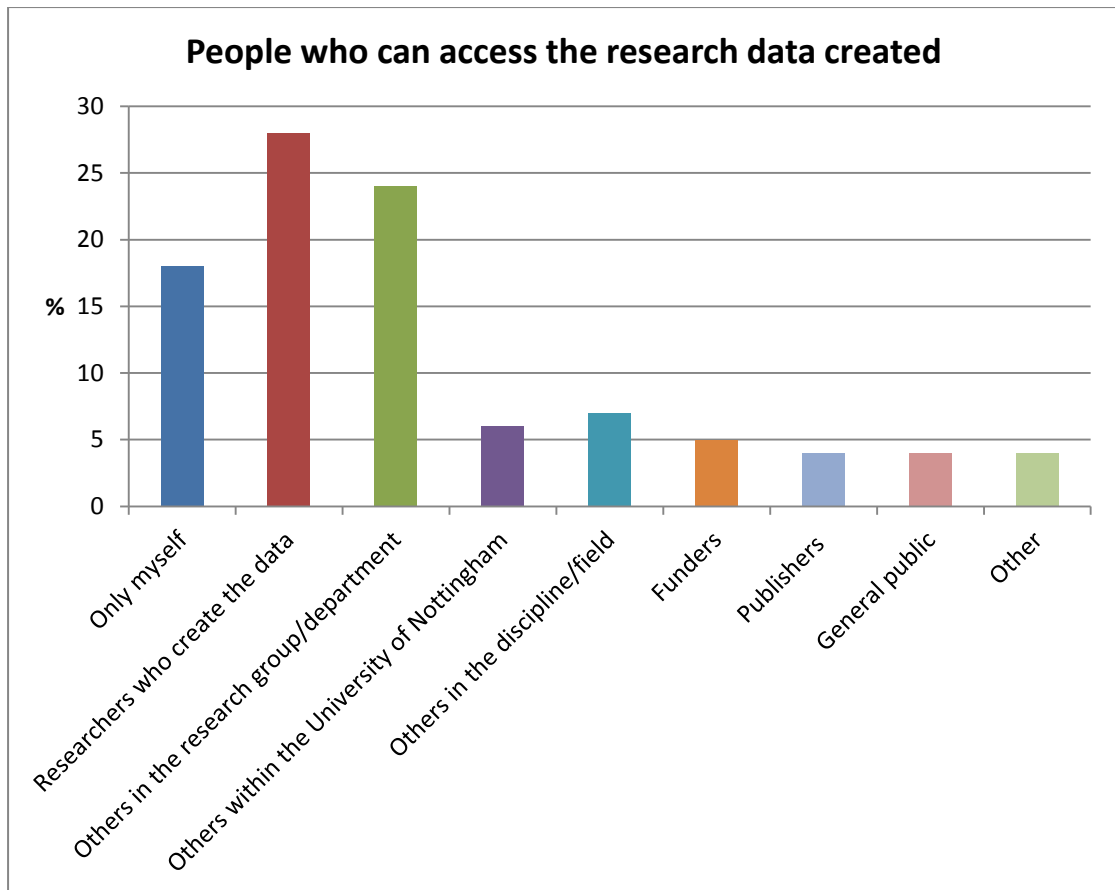
Overall there was a positive response to receiving a form of RDM training; while others questioned whether staff would have time and whether these new processes would be adhered to.

Finally one respondent suggested that the new push to share data might open up new opportunities to generate research revenue. This is not a typical training request, but does highlight that reuse and sharing are seen as important by the researchers themselves.

The overall feeling from responses was one of quite diverse needs in terms of training. Evidently such training would need to be relevant to those needs if it was to be seen as worthwhile, perhaps questioning whether generic or high-level RDM training is sufficient over a subject-based approach.

### 3.10. Accessibility of research data

Question 15 asked respondents who could access the research data they created, this was to ascertain the current level of sharing and outline typical working practices regarding data access. Figure 18 shows the responses.



**Figure 18: People who can access the research data created**

This was a multiple choice/multiple answer question so respondents could have ticked several response options. What is clear from Figure 18 is that the research data created by most respondents is only accessible to people within their research group or department, and for some, it was only accessible to the researcher themselves.

Many respondents elaborated on their responses and explained that their research was at an early stage and therefore there was no data to share. Others (a PhD researcher, a post-doctoral researcher and a Research Fellow) mentioned that they shared their data with their supervisors.

Some said that the data would be more widely accessible once published, while for others, sharing was much more on an ad hoc basis: *"There is no consistent access plan. It depends a bit on where the data is stored"*, while others suggested it depends on the type of research and the types of data being generated. Only one respondent used this section to express a desire to be able to share their data more easily.

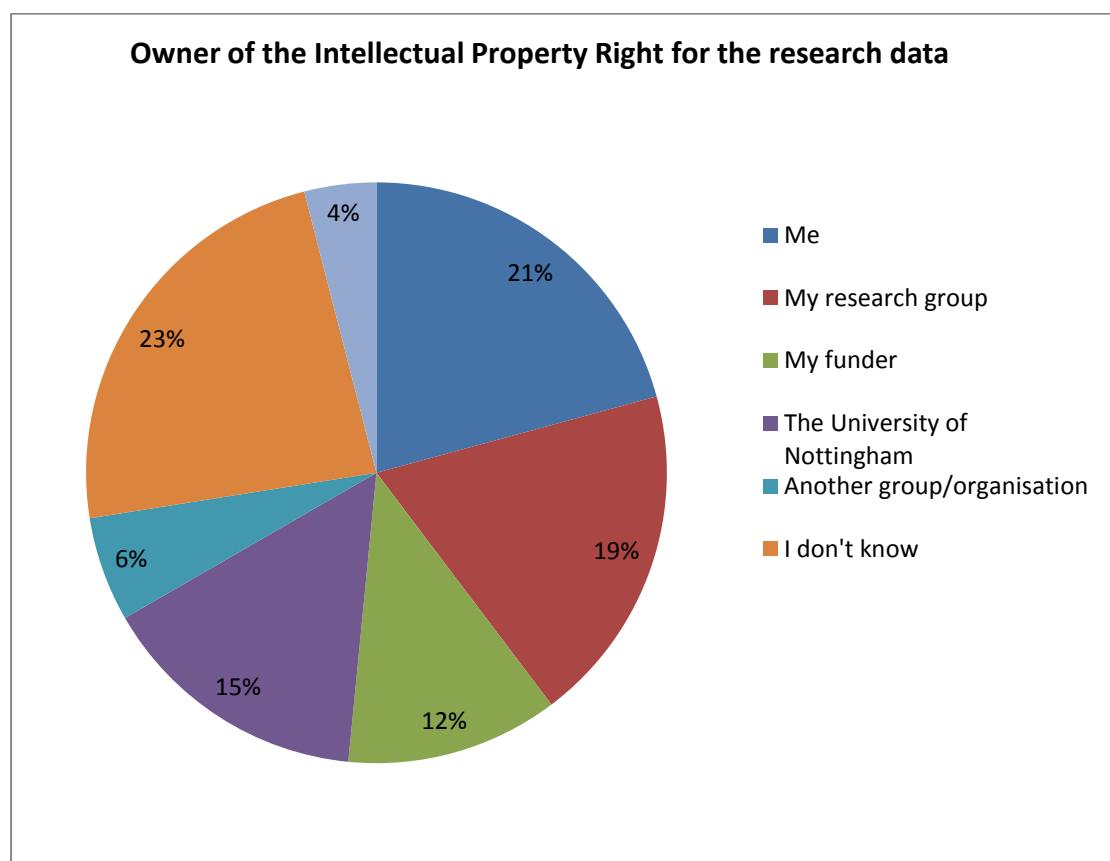
*"Currently funders can only access what we send them but I would like to be able to have a system to be able to share data with sponsors."*



Finally, it was pointed out by one respondent that the degree to which they shared their data was confined by ethics and governance permissions. This is an important area and the following set of survey questions were designed to assess the level of understanding regarding data ownership and the ability to share data, if required.

### 3.11. Intellectual Property Rights (IPR) for the research data

Question 16 asked about the IPR for the research data. Figure 19 shows the percentage of respondents who responded to each category:



**Figure 19: Owner of the Intellectual Property Right for the research data**

The feeling gained from many answers to this question was that it was a very complex issue and that IPR varied from one project to the next.

*"Depends per data resource, we have many."*

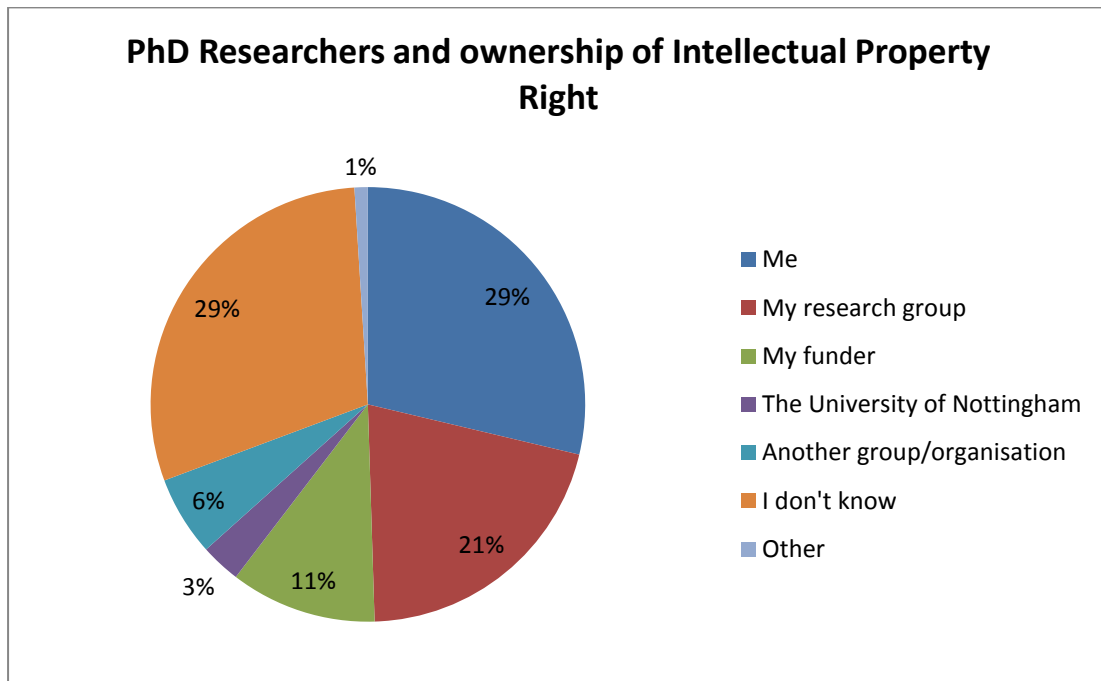
*"In most cases this will be the University though this will vary on a case to case basis."*

*"Various - depends on the data - this is true for pretty much all data and some will be ambiguous."*

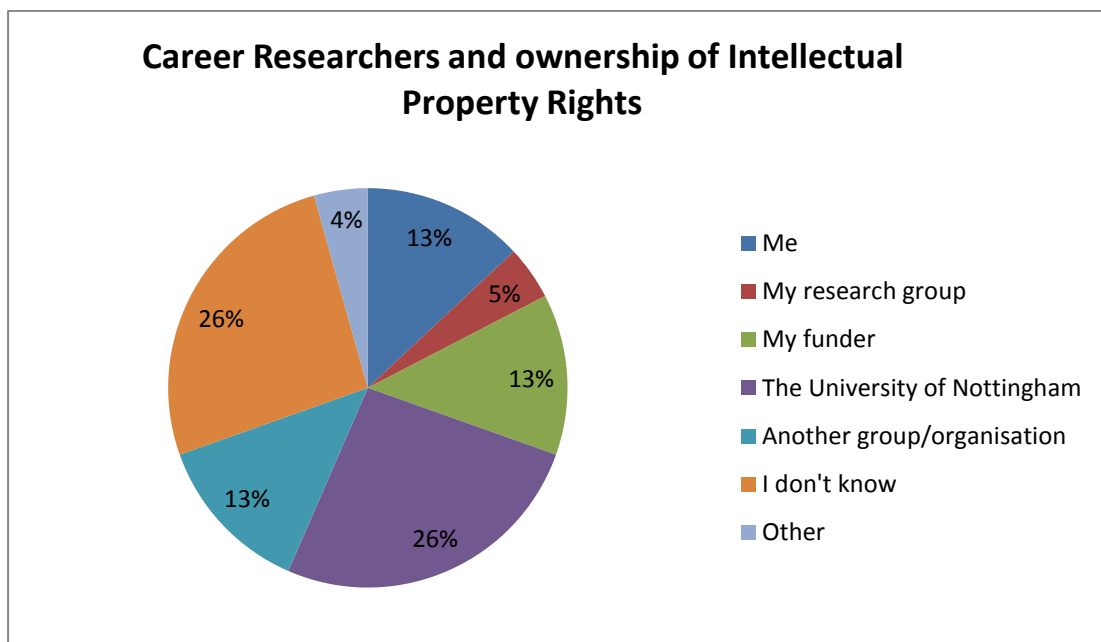
Some respondents seemed to believe that the question was designed to test their knowledge and the question was answered quite tentatively.

*"Is this question designed to catch me out? I think it's a combination of me and the University."*

The comments and spread of answers clearly indicate that IPR is a confusing area and one that could be addressed via training and must be stated in a project's DMP. It was decided to examine the data more closely to establish whether or not responses to the question on who had the Intellectual Property Right of the research data, changed according to the role held by the respondent:



**Figure 20: PhD Researchers and ownership of Intellectual Property Right**



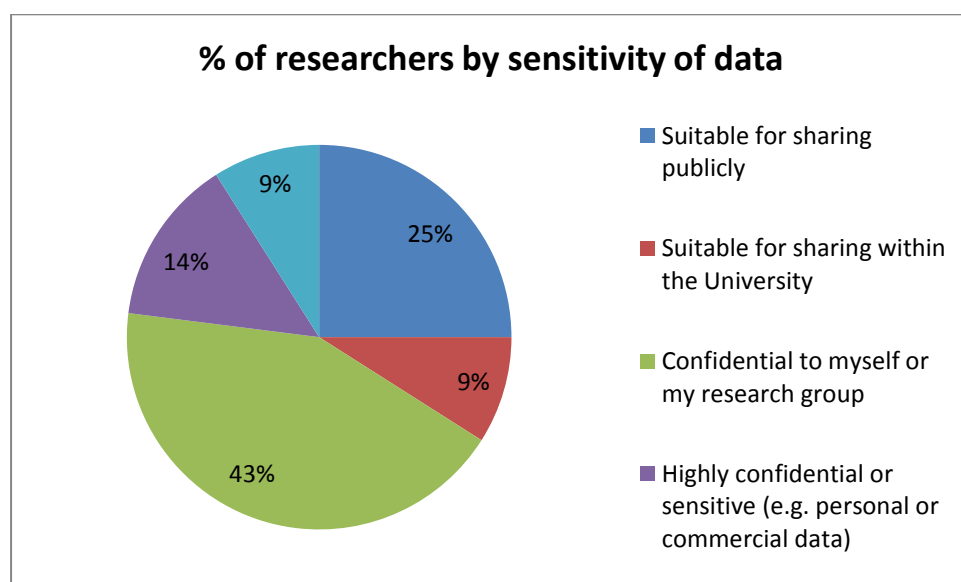
**Figure 21: Career Researchers and ownership of Intellectual Property Right**

Figure 20 and Figure 21 display clear differences in perception of ownership, with very few PhD researchers believing that the University owned the IPR – whereas in reality, the opposite is true in the majority of cases. Further work is required to understand the policies and contracts that influence IPR as this ultimately, will affect whether the data can be shared or not.

### 3.12. Sensitivity of data and suitability for sharing

Respondents were asked about the sensitivity of their research data and how suitable it was for sharing. Figure 22 shows how respondents judged the sensitivity of their data.

Only 25% of respondents answered that their data was suitable for sharing with the general public. Another 9% said their data was suitable for sharing within the university and 43% of respondents answered that their data was either highly confidential or confidential to themselves or their research group.



**Figure 22: % of researchers by sensitivity of data**

Of interest was that 9% of respondents chose to put their answer in the 'Other' category and all except one of those qualified their choice with comments. These comments show that, as in the previous question on IPR, sharing data is a very complex issue.

*"Some of the data is confidential, but other data would fall under highly confidential data"*

*"It is produced during commercial contracts where the client has requested confidentiality and access restricted to themselves and only the person undertaking the work."*

*"I have datasets from multiple projects, with differing requirements. One rather large set (50G and growing) is confidential due to licensing agreements with the provider. Datasets from other projects have to be confidential or suitable for sharing publicly depending on the sample being imaged - however the sheer volume of data (multiple Tb) could be a barrier!"*

*"Raw video data show people receiving and providing healthcare. Clearly this is highly personal data. Ethics permissions involve very restricted use and access to the data[.]"*

As well as the issues mentioned above of commercial interests, size of data and ethical reasons, many respondents said that they would not share their data until they had published from it, due to competition in the field:

*"Highly sensitive academic material from a competitive standpoint which is not shared until published."*

*"The majority is confidential to avoid competition in the field."*

*"It would be better not to share data until they are published."*

This raises the importance of allowing an embargo period between creation of the data and its release and clearly indicates that without such a policy, academics may not engage with RDM. Some respondents also pointed out that while the data was not confidential and therefore could be shared; there were practical issues that would make the reuse of the data extremely difficult.

*"A lot can be shared but is in private formats and won't make much sense!"*

*"It's not necessarily in a format that would be useful to others. There are conventions that would need to be met in order to share it. e.g. to do with editorial standards and making it intelligible and useful[.]"*

*"The data are not confidential and in that sense are suitable for sharing but most data sets are not appropriately annotated for sharing."*

With one respondent stating that without processing the data is meaningless:

*"The data we generate is typically not suitable for sharing without processing- this is why we publish."*

Finally, comments from a couple of respondents suggested that they had felt restricted by the categories provided and had therefore put their answers into the 'Other' category.

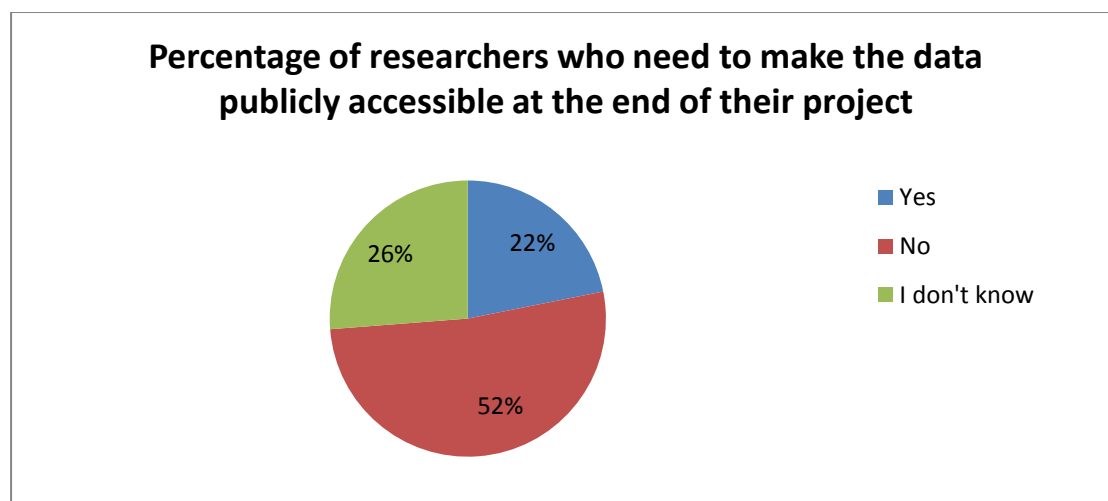
*"You missed the 'other researchers' option ;) It is suitable for academic use, not to the general public."*

"All the above categories are true. We have strict rules for releasing data into each category."

The survey then followed up this topic and examined awareness of sharing requirements.

### 3.13. Making research data publicly available

Question 17 asked respondents whether or not they needed to make their data publicly accessible at the end of their project:

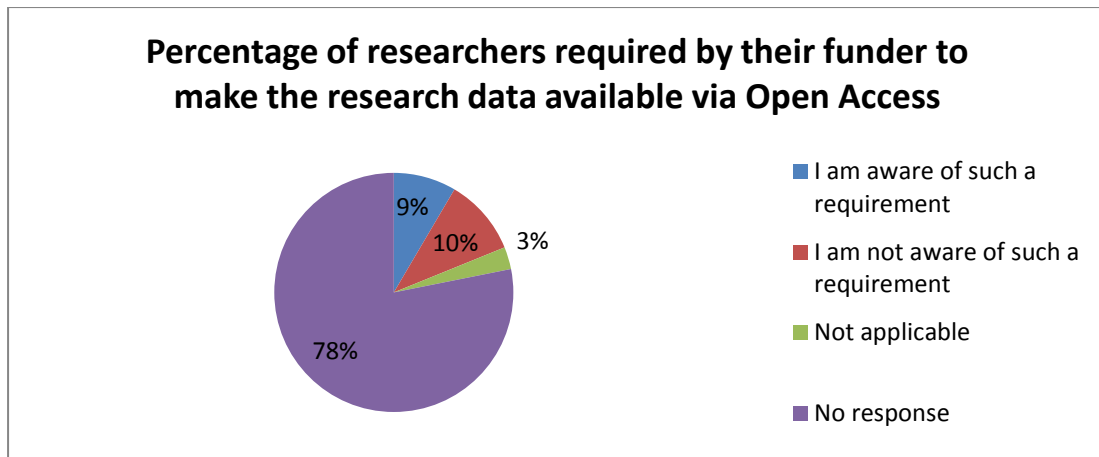


**Figure 23: Percentage of researchers who need to make the data publicly accessible at the end of their project**

Figure 23 shows that only 22% of respondents were required to do this. However, 26% didn't know either way. Question 17 then concluded with a follow-on optional question (17a), which asked if they need to make their research data available at the end of the project via Open Access<sup>14</sup>; seeking to ascertain if *their funder* was the one requiring them to make their research data publically accessible. Figure 24, illustrates that out of the 22% who said "yes" in the first part of question 17, only 9% of these said that they were aware of such a requirement from their funder.

---

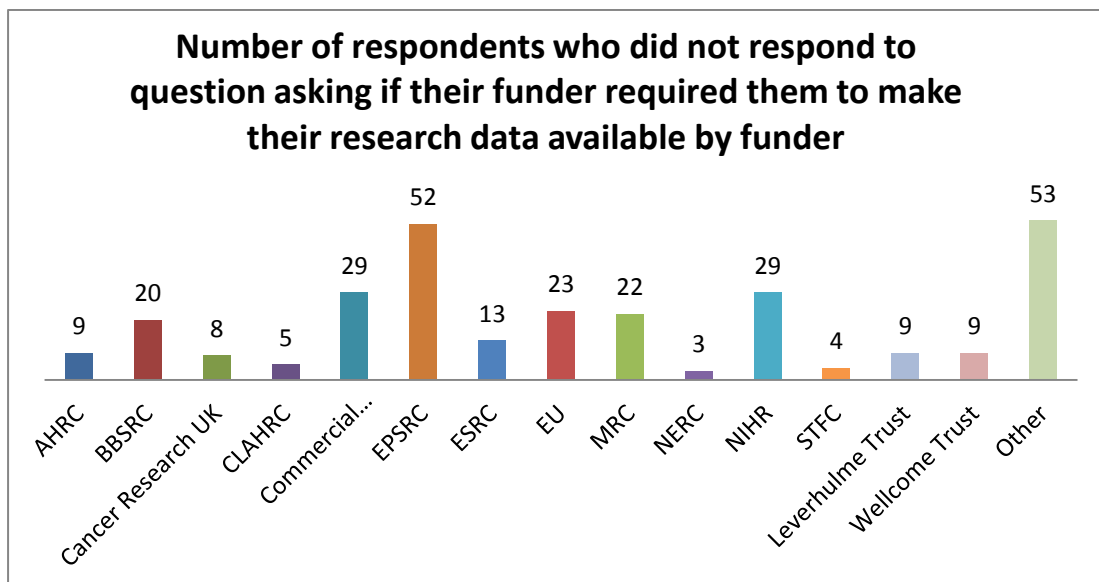
<sup>14</sup> Since analysing the survey, we have questioned whether we should have used the wording "Open Access" (OA) in Question 17a. We do recognise that the two areas of OA and data sharing are distinct and to mix these concepts may have been confusing to some. However, the concepts are frequently associated together under the guise of "research outputs", so we believe that the gist of the question was still clear and hence the results are valid. In retrospect, we believe it would have been better to reword the question and use "If yes, does your funder require you to share and make your research data freely available?"



**Figure 24: Percentage of researchers required by their funder to make the research data available**

This is an interesting finding, as although Figure 23 demonstrates a low awareness of requirements to share data publically, it also indicates that there are researchers who are sharing their data independently of any funding mandates.

A more detailed analysis revealed that respondents, who are funded by the University, are likely to be taking note of the University of Nottingham's Code of Research Conduct and Research Ethics<sup>15</sup> that states that people should share their data. More interesting though, is that two thirds of those who choose not to answer 17a are externally funded and in all likelihood, *should* be sharing their data. A detailed analysis of their funding sources is shown in Figure 25.



**Figure 25: Number of respondents who did not respond to question asking if their funder required them to make their research data available by funder**

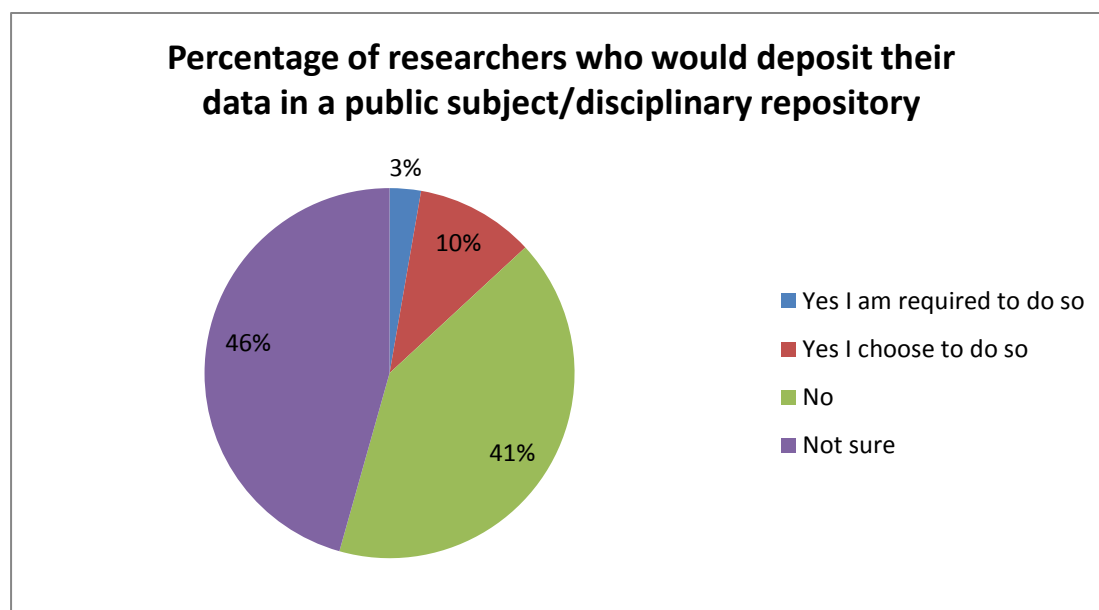
<sup>15</sup> <http://www.nottingham.ac.uk/fabs/rgs/documents/code-of-research-conduct-and-research-ethics-approved-january-2010.pdf>

Clearly, the major RCUK funders are funding these researchers, yet the awareness amongst researchers of their requirements for data sharing is low. It is unclear whether this lack of awareness is due to the relatively recent changes in policy, although BBSRC for example, have had a data sharing policy in place since 2007. Another answer could be that these policies have simply not been communicated to the researchers – either via the funding bodies themselves or via the University. A brief analysis of awareness based upon Faculty and role type, revealed that the MHS and Science Faculties had the highest awareness. “Research Fellows” and the “Lecturer/Researcher” roles showed highest awareness across all role types, suggesting that PIs from the Physical and Medical Health Sciences are again taking the lead.

Although awareness is low, there are good resources available to share with researchers (i.e. the DCC funding policies table<sup>16</sup>) and a strong argument for a programme of RDM advocacy events at the very least.

### 3.14. Depositing data in a public subject/disciplinary repository

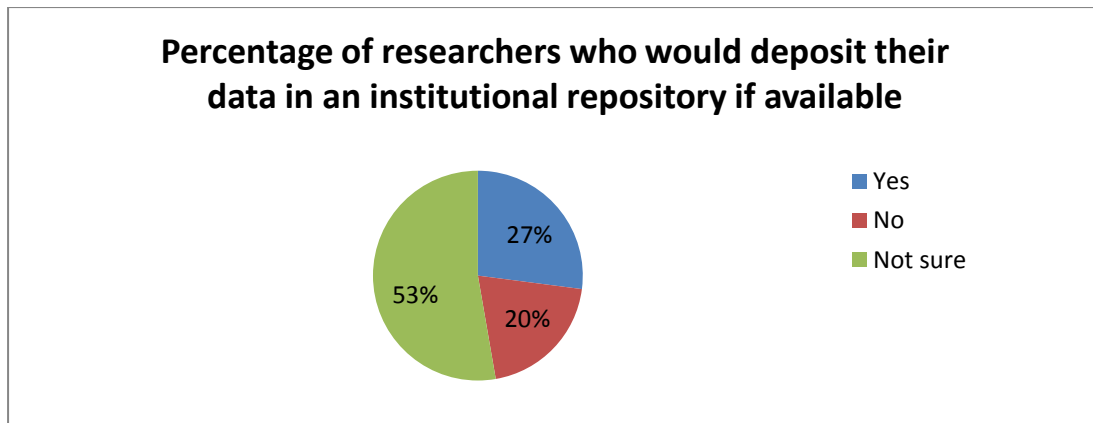
As can be seen from Figure 26, when respondents were asked whether or not they would deposit their research data in a public subject/disciplinary repository, only 13% answered that they would. Only 3% did so because they were required to do so, and another 10% said they chose to of their own volition. 41% said they would not do so, but no reasons for this were collected.



**Figure 26: Percentage of researchers who would deposit their data in a public subject/disciplinary repository**

<sup>16</sup> <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>

Respondents were then asked if they would deposit their data in an institutional repository if that was available. As Figure 27 shows, 27% answered that they would, while 20% said that they would not.



**Figure 27: Percentage of researchers who would deposit their data in an institutional repository if available**

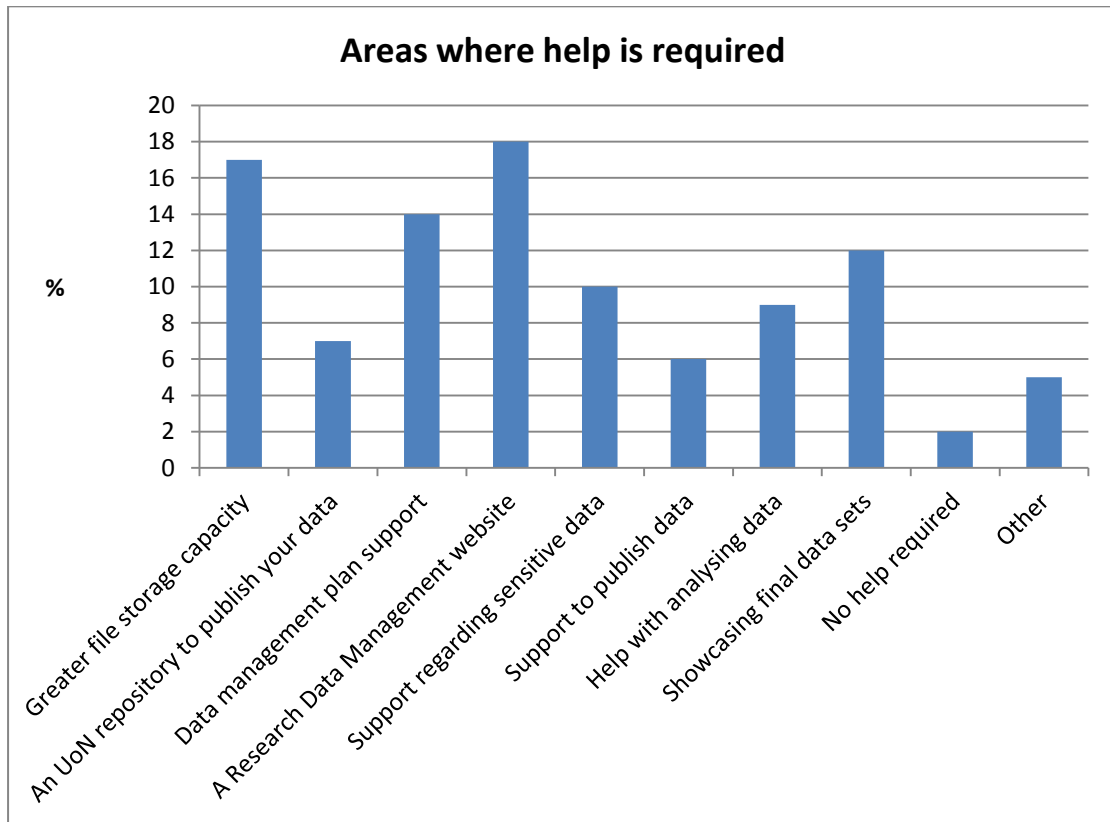
Thus a more positive response was obtained for an institutional repository than for a public subject/disciplinary one. Again, the reasons for respondents' choices were not collected and this is well worth following up in later research, particularly to understand the 53% of researchers who are "Not sure" of whether to deposit data in an institutional repository.

Follow-up interviews revealed that researchers have concerns over privacy of data and this may be the primary cause of the low response to using an institutional repository. Researchers noted that repositories are usually associated with the open sharing of research outputs. Therefore, in terms of providing a repository to store research data, there's a strong requirement for a repository to both be able to share and secure data as appropriate.



### 3.15. Areas where help is required

The final question on the survey asked respondents to select areas where they would like to receive help. The results are shown in Figure 28 below:



**Figure 28: Areas where help is required**

The areas which most people requested are:

1. A Research Data Management website for guidance and support
2. Greater data storage capacity
3. Data management support when writing a research proposal
4. Help to make better use of your final data sets (e.g. create website to showcase data)
5. Support regarding sensitive data

A huge variety of answers was provided in the 'Other' section. These included:

- The provision of a secure and reliable off-site back up facility
- A School-wide scheme for archiving data
- A central shared tape drive for storage of data

- Help with exchanging data with other collaborators at different institutions
- Training and support with IT tools for user management and database construction/maintenance
- Advice, help and training in dealing with data management issues in large database data

The majority of the comments concerned the technical management of data, such as backup and databases and perhaps indicates where the researcher's priorities lie at present – certainly more on the side of managing and working with data, rather than complying with funding or Institutional mandates.

Other areas that were raised in comments included the funding source of RDM activities and the volume of data that will be stored long-term – both areas that require significant work and that are unclear at present.

## 4. Conclusions

The response rate of the survey was very positive and allows valid conclusions to be made. Notable observations are:

- The diversity of data types and the strong presence of non-digital data such as lab notebooks.
- There are multiple locations for the data and therefore, the ad-hoc strategies of back-up.
- Encouragingly there are very few researchers who do not back-up at least a sub-section of their data.
- As expected, University storage is widely used but media such as external hard drives and USB sticks are also popular, as are external services such as DropBox.
- The range of data sizes means the standard University provision of 4 GB of space may be insufficient. The new offering of 1TB of research data file storage will meet this gap, but may be overly generous in the majority of cases.
- Researchers favour lone or collaborative working, with few allowing access to their data for those outside of the University or their research group.
- Sensitive data, IPR rights and the sense of ownership to the data will doubtless, hamper efforts to share data. Overall the responses indicate that certain areas such as the medical fields will require additional effort to investigate if and how sensitive data can be shared.
- Training appears to be high on the agenda for many, with very few expressing no interest at all. Key areas included help with DMPs, metadata, storing data and funding body requirements sessions.
- The funding analysis revealed a surprisingly low awareness of funding requirements regarding data sharing.

The results show that the current level of RDM awareness and training is in its infancy. Therefore the timescales for engaging with all researchers is a long-term commitment beyond the life time of the ADMIRe project. Although it remains to be seen the level of service that will eventually be required, a high proportion of researchers would value a dedicated RDM website, but whether this is enough to satisfy the majority of RDM enquiries is unclear.

Therefore the University must take a pro-active stance both in terms of policy and monitoring compliance to RDM; responses indicate that training, an RDM support service and improved storage and backup facilities would have a good uptake if they were introduced at an institutional level. Less encouraging is the attitude towards sharing data – again suggesting that the benefits to the researchers of publishing data need to be articulated clearly at all stages.

Overall a strong case can be made to support researchers who are faced with the challenge of developing good RDM practices; certainly there are clear requirements for advocacy, training and a revised technical infrastructure to allow them to meet funding and University policies. Although, as the JISC funding for this project draws to a close, it remains to be seen who will develop these findings into a coherent and relevant service.

## 5. Appendix

### 5.1. RDM Survey questions and accompanying text

#### Aim of this research data management survey

Welcome and thank you for agreeing to take part in The University of Nottingham's research data management survey. The survey is intended for University of Nottingham researchers across all Faculties.

This survey is conducted by Information Services, who along with colleagues in Research Graduate Services, the library and the Management Board, are developing a research data management infrastructure to support the research data lifecycle. This project is part-funded by JISC and is endorsed at the highest level across the University.

This part of the project will discover how research data is used and managed across the University.

The questionnaire is designed to:

- Assist the project team to understand the data held by the researchers
- Discover the influences and barriers to managing research data
- Establish what advice and support you require
- Identify current levels of research data management practice in faculties.

We will use the information you provide to:

- Inform a series of research data pilot studies
- Inform project requirements gathering
- Assess what data we are seeking to manage
- Deliver tools, infrastructure, and policies to facilitate good research data management practice at The University of Nottingham
- Feedback good practice to other institutions.

*This survey can be completed anonymously but if you would like to be entered into the prize draw to win an iPod shuffle, then please enter your name and email address at the start of the survey.*

#### Instructions

The survey consists of 20 questions and should take 15-20 minutes to complete.

## Personal details

If you wish to be entered into the prize draw for an iPod shuffle then please answer questions one and two (optional questions). Information entered here will not be used for any other purpose.

1. Name (optional)
2. Email address (optional)
3. Faculty
  - o Arts
  - o Engineering
  - o Medicine and Health Sciences
  - o Science
  - o Social Sciences

Textbox:

4. Which of the following best describes your research role:
  - o PhD researcher
  - o Post-doctoral research
  - o Research fellow
  - o Lecturer/Researcher
  - o Career researcher
  - o Other:

## About your research data

In this section we would like to find out about your current research and how you create and manage your research data.

5. What types of research data do you create or work with as part of your research?  
*Select all that apply:*
  - o Documents (text, PDF, Microsoft Word)
  - o Spread sheet (e.g. Excel)
  - o Websites
  - o Notebooks/diaries
  - o Databases (e.g. Access, MySQL, Oracle)
  - o Questionnaires, transcripts, codebooks
  - o Audiotapes, videotapes
  - o Photographs, films
  - o Slides, artefacts, specimens, samples
  - o Collection of digital objects acquired and generated during the process of research
  - o Raw data files generated by software, sensors or instruments files
  - o Models, algorithms, scripts
  - o Contents of an application (input, output, logfiles for analysis software, simulation software, schemas)
  - o Other

6. Where is this research data stored? *Select all that apply:*

- Hard disk drive of campus computer
- Hard disk drive of off-campus computer
- Hard disk drive of laptop/netbook
- Hard disk drive of instrument/sensor which generates data
- External hard drive
- USB/Flash drive
- Shared drive/university server
- An institutional repository (please specify in 'Other')
- Web based service, e.g. Dropbox, Flickr, Google Docs (please specify in 'Other')
- CD/DVD
- Email client/server
- VHS/Video Cassette
- Floppy Disk
- Audio Cassette Tape
- Photographs
- Slides
- Microfiche
- On paper
- Other: (please )

7. Please estimate the volume of research data across all of your work:

- <1 GB
- 1-50 GB
- 50-100 GB
- 100-500 GB
- 500GB-1 TB
- 1-50 TB
- 50-100 TB
- >100 TBs
- I don't know

8. How frequently is your research data backed up?

- Daily
- Weekly
- Ad-hoc
- Don't know
- Never

If yes, where is it backed up?

- UoN file storage
- External hard drive
- USB/memory stick
- CDs/DVDs
- Server managed by IS
- Server managed by yourself or research area

- External e.g. Dropbox, Google Docs, Amazon S3  
(please specify in 'Other')
- Don't know
- Other:

9. Do you document or record any metadata about your data? This is to make data more meaningful or easier to search for.
- Yes
  - No
  - I don't know

If yes, do you use any standards or guidelines?

- Yes (please specify)
- No
- I don't know

## Research Data Management training and requirements

10. Are you currently working on a funded research project?

- Yes
- No

Who are the funders of your research? (select all that apply)

- AHRC
- BBSRC
- ESRC
- EPSRC
- MRC
- NERC
- STFC
- Cancer Research UK
- Wellcome Trust
- The Leverhulme Trust
- Commercial organisations (please specify)

11. Have you developed a research data management plan for your project?

- Yes
- No
- I don't know

12. Are you aware of any policy or requirements from your funder regarding research data management?

- Yes
- No
- Not applicable



13. Have you ever received any research data management training?

- Yes
- No

If yes, please provide further details:

14. Would you like to receive data management training in the following areas? *(select all that apply)*

- Developing a research data management plan
- Documenting your data
- Formatting your data
- Storing your data
- Sharing your data
- Creating metadata for data
- Ethics and consent
- Funders requirements and research data management
- Copyright and intellectual property right (IPR)
- Data repositories and Open Access
- Other *(please specify)*

## Sharing your data

Ways of sharing your research data include data repositories (subject and institutional), data banks and data centres, submission to a journal to support publication, and informally between researchers.

15. Who can typically access the research data you are creating? *(select all that apply)*

- Researchers who help the create the data
- Others in the research group/department
- Others within the University of Nottingham
- Others in the discipline/field
- Funders/publishers
- General public

*Other (please specify):*

16. Who has the Intellectual Property Right for your research data?

- Me
- My research group
- My funder
- Another group/organisation
- I don't know

Other (please specify)

17. At the end of your project do you need to make your data publicly discoverable and accessible via Open Access?

- Yes
- No
- Partially
- I don't know

If yes, does your funder require you to make your research data available via Open Access (OA)?

- I am aware of such a requirement
- I am not aware of such a requirement
- Not applicable

18. Would you deposit your data in a public subject/disciplinary repository, e.g. arXiv, Visual Arts Data Service?

- Yes, I am required to do so
- Yes, I choose to do so
- No
- Not sure

19. Would you deposit your data in an institutional repository if available?

- Yes
- No
- Not sure

## Support for Research Data Management

20. The University Of Nottingham is committed to supporting researchers across the research lifecycle. We would like to know where you require help, please select all that apply:

- Greater file storage capacity
- An UoN repository to publish your data
- Data management support when writing a research proposal
- A Research Data Management website for guidance and support
- Support to publish data to external subject repositories
- Help with analysing data
- Help to make better use of your final data sets e.g. create a website to showcase your data

Other (please specify)

## Final page of the survey

### Data Protection Statement

*All your comments will be treated as confidential and anonymised information will be included in our project reports.*

*Any personal and identifiable data we collect from your survey responses will be accessible to the project team members (named at the end of this survey) only.*

*Your data may be used to inform and develop research data management tools, infrastructure, and policies at the University of Nottingham.*

Thank you for completing the research data management survey. We appreciate the time you have taken. Your responses will help us understand how research data is managed at The University of Nottingham. The service website is due to be launched in October 2012 within the main Research section: <http://www.nottingham.ac.uk/research/>

In the meantime, should you require more information on the service or if you can assist us, please contact:

**Research Data Management Project Manager:** Dr Tom Parsons  
[tom.parsons@nottingham.ac.uk](mailto:tom.parsons@nottingham.ac.uk)

**Research Data Management Service Developer:** Laurian Williamson  
[laurian.williamson@nottingham.ac.uk](mailto:laurian.williamson@nottingham.ac.uk)

For more information on how JISC, ourselves and other universities are helping to manage and share data, please view:

The University of Nottingham JISC ADMIRe blog:  
<http://admire.jiscinvolve.org/wp/>

The JISC Research Data Management Infrastructure Project:

[http://www.jisc.ac.uk/whatwedo/programmes/di\\_researchmanagement/managingresearchdata/infrastructure.aspx](http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata/infrastructure.aspx)

### Acknowledgements

Surveys produced by projects which are part of the JISC Managing Research Data Programme (JISCMRD), including those produced by UWE, Newcastle University, University of Lincoln and the University of Exeter. Material produced by CARDIO Pulse Check and the Digital Curation Centre (DCC).