Research Data Management Project

A DAF investigation of research data management practices at The University of Northampton

Report prepared by:

Edward Alexogiannopoulos and Sam McKenney
Project Researchers

Miggie Pickton
Research Support Specialist, Information Services

September 2010

Table of Contents

Acknowledgements

This project could not have been undertaken without the help of the two project researchers, Sam McKenney and Edward Alexogiannopoulos, appointed through the Graduate Boost[1] programme. Graduate Boost is designed to provide unemployed graduates with work experience and postgraduate credits towards an MBA qualification. It is managed by 3e, a social enterprise employment agency, on behalf of the Northampton Business School, and is funded by the Higher Education Funding Council for England (HEFCE) and the European Union European Social Fund. We are grateful to all of these sponsors.

We also acknowledge the training and support provided by colleagues as part of the KeepIt project. KeepIt is a JISC funded collaborative project involving the Universities of Northampton, Southampton and the University of the Arts in London. The KeepIt training course introduced us to a variety of tools for managing digital preservation in repositories; the Data Asset Framework used here was one such tool. We are grateful to the Digital Curation Centre for developing and sharing the Data Asset Framework[2].

Last but not least, we greatly appreciate the support for the project that we received from the research community at The University of Northampton. Researchers and research managers across the institution were generous with both time and advice to ensure we gained the best possible overview of research data management practices at the university.

Thank you.

[1] http://graduateboost.co.uk/grad-home.htm
[2] http://www.dcc.ac.uk/resources/tools-and-applications/data-asset-framework

Executive Summary

This report describes the University of Northampton Research Data Project, conducted over a period of eight weeks in Summer 2010.

Following the tried and tested Data Asset Framework (DAF) methodology, the project sought to investigate the types of data held by researchers throughout the university, researchers' existing data management practices, and the risks associated with these practices. It aimed to provide evidence to inform a possible new data management policy and services to satisfy the requirements of researchers and funders. Finally, it hoped to raise awareness amongst researchers about good data management practice, including the provision of short and long term data storage and access.

Drawing on the experience of previous DAF implementations, the project used two methods for gathering data: an online survey which attracted 80 respondents, and in depth interviews with 16 researchers. The survey covered a wide range of issues including the types, sizes and formats of research data held; its ownership; means of storage; security arrangements; sharing and access over the short and long term; and the requirements of funders. The interviews enabled the project team to follow up key findings from the survey and gather additional technical information on specific data objects.

A number of themes emerged. Three generic types of researcher were identified, based on their demonstrated different needs and behaviours with respect to research data: the research student, the independent researcher and the group researcher/collaborator.

Some common behaviours were identified, for example, researchers overwhelmingly use Microsoft software for creating documents and spreadsheets and so habitually create .doc and .xls file types; similarly, .jpeg is the preferred format for image files. In contrast, there is much greater variation in the file types used for databases, audio and video files. These findings have significant implications for preservation planning.

Data storage needs and behaviours vary throughout the research lifecycle, with different storage devices being prominent at the data collection, analysis and project completion stages. For those that need to share data, a shared server is effective, but where this is not available, email is most frequently used.

Very few Northampton researchers have applied for funding from a body that mandates open access to research data and only just over half are interested in a university repository for data (either for open or closed access).

Several problems and concerns were raised: there is uncertainty surrounding the ownership of data; data are still being collected in out-dated formats; data management practices are guided by intuition rather than informed by good practice; data are sometimes neglected once a project is complete; the university's shared server space is under-exploited; and researchers are sometimes ill informed, or even misinformed, of the services available to them.

Potential solutions centre on the creation of appropriate policy on research data management with advocacy, guidance, training and documentation to support this.

Nine recommendations are made:

1. This report to be presented to senior research managers and to the university's Research Committee for discussion.

2. A university research data policy to be drafted and approved by Research Committee.

3. The university to clarify its position on the ownership of research data and other research outputs generated by staff and research students at Northampton.

4. Information Services, in conjunction with the Graduate School, to develop and promote training sessions on 'Data management for researchers'. Based on existing records management training, these will focus on the specific needs of researchers.

5. Information Services to produce a research life cycle based guide to research data management.

6. Information Services to further develop and disseminate expertise in preservation planning to support researchers wishing to store and access their data over the medium to long term.

7. The Research Support Specialist to present the findings of the report to Schools and Research Centres, together with advice and guidance in line with the new research data policy and information concerning the ownership and exercise of rights to research data.

8. This report to be deposited in NECTAR.

9. The Data Curation Centre to be informed of the project and invited to create a link to NECTAR from the DAF website so that other institutions may read about Northampton's implementation of the framework.

1. Introduction

This report documents the purpose, planning, implementation and results of the Research Data Project conducted within the University of Northampton (UoN) in 2010.

The project was launched on 10th May 2010 by the Department of Information Services (IS).

The project team comprised Project Manager Miggie Pickton (Research Support Specialist), two Project Researchers, Sam Mckenney and Edward Alexogiannopoulos, and a Project Board consisting additionally of Phil Oakman (University Records Manager) and Philip Thornborow (Collections and Learning Resources Manager).

1.1 Rationale for project

The collection, analysis and storage of data have always been core to the research process. However, with the increase in 'Big Science', the proliferation of multi-disciplinary research projects and rapid changes in technology, the scale and complexity of these have grown enormously. Combine this with increasing pressure from the UK government for research outputs, including data, to be made openly accessible, a political and social climate that demands accountability, and a legal framework that permits individuals to make Freedom of Information (FOI) requests for research data, then it is clear that there is a great need for a robust approach to managing research data.

Set in this context, the motivating factors behind the project at Northampton were as follows:

- At the start of the project no university-wide data management or storage policy or procedure existed.
- Researchers at Northampton were beginning to win funding from organizations that either required or requested that research data should be made accessible to the wider community.
- IS held no comprehensive picture of how researchers create, use, store or share data at the university. This information was needed to inform the department's support role, particularly with regard to providing new services.
- In NECTAR, the university's open access repository, there was sufficient capacity to store research data, should this be required by the research community.

1.2 Aims

1. To investigate the types of data held by researchers throughout the university, researchers' existing data management practices and the risks associated with these practices.
2. To provide evidence to inform a possible new data management policy and services to satisfy the requirements of researchers and funders.
3. To raise awareness amongst researchers about good data management practice, including the provision of short and long term data storage and access.

1.3 Stakeholders

The primary stakeholders of the project, as well as their roles and interests are as follows:

	Role	**Interest in Project**
Researchers	Undertaking surveys and interviews	New services to help researchers store, backup, collaborate and provide access to data may result. Potential creation of a UoN policy and services that will facilitate bids for future funding and fulfil funding requirements.
Research managers	Receive report, contribute to and implement new policies and procedures	Gain evidence for fulfilment of research requirements at the university. Enhancement of the university's research environment through potential new policies and services
Information Services	Conducting project and producing new policy/services/training	Gaining an understanding of data management and how to provide better academic support. Building relationships with the research community.
University of Northampton	Funding new services/training	Efficiency and quality of UoN research may increase along with a corresponding gain in reputation/external funding.

2. The Data Asset Framework (DAF)

The Data Asset Framework (DAF) was selected as the methodology for this project. DAF has been successfully used in other universities and has been shown to yield useful results.

DAF is a framework methodology for assessing data management and holdings in an institution. Initially known as the Data Audit Framework, the development of DAF was funded by the Joint Information Systems Committee (JISC) following a recommendation from the report 'Dealing with Data'[3]. As a framework, it is a general procedure and institutions are encouraged to adapt its methodology as required. It has four stages which are set out in the DAF Implementation Guide[4] as follows:

Stage 1 is for planning, defining the purpose and scope of the survey and conducting preliminary research.

Stage 2 is about identifying what data assets exist and classifying them to determine where to focus efforts for more in-depth analysis.

Stage 3 is where the information life cycle is considered to understand researchers' workflows and identify weaknesses in data creation and curation practices.

Stage 4 pulls together the information collected and provides recommendations for improving data management.

Definition:

In this report, the term 'data object' is used to denote a quantum of data. A single data object consists of all homogenous data, documentation and metadata relating to a specific project. For instance a one-off interview transcript would be a single discreet data object as would a collection of a thousand zoological photographs and corresponding documentation.

---

[3] Lyon, L. (2007) *Dealing with data: Roles, rights, responsibilities and relationships* [online]. Available from: http://www.jisc.ac.uk/whatwedo/programmes/digitalrepositories2005/dealingwithdata.aspx [Accessed 17th September 2010].

[4] Digital Curation Centre (2009) *Data Asset Framework: Implementation guide* [online]. Available from: http://www.data-audit.eu/docs/DAF_Implementation_Guide.pdf [Accessed 17th September 2010].

2.1 Previous DAF Implementations

Stage 1 of this project included a desk-based investigation of previous implementations of DAF. The wealth of DAF implementation reports available openly online is commendable, none more so than 'DAF Lessons Learned'[5] . The reports of individual universities also deserve acknowledgement[6][7][8].

The following issues frequently arose in different institutions:

- **Timing** of DAF implementations was crucial at Glasgow and Bath. High participation required timing the project to fit in with fieldwork, ongoing research projects, annual leave and examinations.

- **Time** was a limiting factor for all implementations and prompted the restriction of scope. Edinburgh said that time could have been used more efficiently by producing a comprehensive survey and promptly arranging follow-up interviews. Southampton warned that a great deal of time could be used by rigorous transcription of interviews and suggested a quicker style of note-taking or recording.

- **Advocacy** was important to encourage high participations rates. Glasgow found that gaining personal introductions and obtaining advocacy from key data managers was helpful. Bath recommended 'badgering' senior staff as needed to progress the implementation.

- **Scope and granularity** was a key consideration for all implementations. King's College London found that scope had to be limited to the one department that was willing to offer advocacy for the project. More broadly, institutions found that it was advantageous to be flexible about scope as time constraints and new information during the project necessitated changes.

---

[5] Jones, S. (2008) *Data Audit Framework lessons learned report: GUARD audit* [online]. Available from: http://www.data-audit.eu/docs/DAF_lessons_learned.pdf [Accessed 17th September 2010].
[6] Martinez-Uribe, L. (2008) *Using the Data Audit Framework: An Oxford case study* [online]. Available from: http://www.disc-uk.org/docs/DAF-Oxford.pdf [Accessed 17th September 2010].
[7] Jerrome, N. and Breeze, J. (2009) *Imperial College Data Audit Framework Implementation: Final Report* [online]. Available from: http://ie-repository.jisc.ac.uk/307/ [Accessed 17th September 2010].
[8] Gibbs, H. (2009) *Southampton Data Survey: Our experience and lessons learned* [online]. Available from: http://www.disc-uk.org/docs/SouthamptonDAF.pdf [Accessed 17th September 2010].

- **The Data Seal of Approval**[9] was recommended by Bath[10] as a good baseline to assess services provided as a result of a DAF implementation.

- **Obtaining as much information as possible** was a general principle advocated by most institutions at both the survey and interview stages since researchers' time is limited and should be fully exploited when offered. Oxford suggested particular attention should be given to disentangling ownership and authorship issues concerning data objects.

3. Methodology

3.1 Procedure

The DAF methodology was adapted following the recommendations of previous implementations and to take account of the circumstances of the University of Northampton:

Stage 1: The planning stage occupied week 1 of the project (10th-14th May) and consisted primarily of familiarisation with DAF and relevant projects in UoN (i.e. KeepIt[11] and NECTAR). This stage was concluded by the first Project Board meeting on the 14th May which introduced the methodology as it was then planned and discussed particular issues for investigation such as how to ensure high participation and how to address questions of data access.

Stage 2: The identification stage consisted of three parts: interviews with research leaders, a pilot survey and a 'live' survey (17th May-21st June).

The research leader interviews had two aims: first, to gain a basic understanding of research within the schools (which would inform the construction of the surveys) and, second, to obtain 'buy-in' from the leaders in the hopes that their advocacy would encourage greater participation in the project. The first aim was mostly achieved with five out of the six school research leaders agreeing to interview. However, due to the limited availability of the leaders only one interview actually occurred prior to the piloting of the survey. 'Buy-in' was

[9] http://www.datasealofapproval.org/
[10] Ball, A. (2010) *Review of the state of the art of the digital curation of research data* [online]. Bath: University of Bath. Available from: http://opus.bath.ac.uk/18774/ [Accessed 17th September 2010].
[11] KeepIt is an ongoing JISC-funded project, jointly conducted by the Universities of Southampton and Northampton and by the University of the Arts in London. It is exploring issues surrounding digital preservation, specifically the preservation of repository content. Further details of the project are available on the project website: http://preservation.eprints.org/keepit/

obtained to some degree from all research leaders interviewed and, no doubt, the high survey response rate owes a great deal to this.

The pilot survey was constructed using the Bristol Online Surveys (BOS) tool. The pilot exercise proved to be valuable, despite only five responses, to tailor the questions to the interests and circumstances of UoN's researchers. It also averted disaster by revealing that BOS surveys do not display correctly on Internet Explorer if the text is copied into the building tool whilst still holding MS Word formatting. This mistake was not repeated in the 'live' survey.

A copy of the 'live' survey is included in Appendix 2. It is significantly longer than many surveys used in previous implementations. The survey's length and level of detail were designed with the intention of building up a broad initial picture of research at UoN. This saved time in terms of follow-up interviews, but perhaps came at the cost of deterring some potential survey participants. This cost to the response rate was offset by a number of measures described later.

Stage 3: The assessment stage (7th June – 25th June) took the form of a review of the survey results alongside a campaign of follow-up interviews. The interviews allowed in-depth assessment of the data management problems and interests of UoN's research community. The follow-up interview plan is shown in Appendix 3.

The results of this stage enabled the project researchers to better understand the types of researcher and research at UoN, the common formats of their data, their approaches to storing data and their attitudes to sharing it.

Stage 4: The collating stage (28th June – 2nd July) reviewed the planning, methodology, findings and other documentation produced by the project to make recommendations to inform UoN's research data management practices and policy. The findings are described in this report.

3.2 Scope

At stage 2 the scope of the project was cast as wide as possible:

- The survey was open to every researcher in the university.
- Researchers were asked about all data they hold, regardless of its age.
- Both digital and non-digital (e.g. paper-based questionnaires, VHS video, photographs etc.) data were considered in the survey.

This broad approach was used because:

- Obtaining high response rates has been challenging in other DAF implementations and many results are required to give an accurate picture of data management in the university.

- It was important to develop a rich and representative understanding of research data at the university.
- By restricting the project to digitally stored data, policy creation might have been simplified and the time-consuming process of conversion and transcription could have been ignored. However, it is non-digital data which are more liable to degradation, poor documentation and inadequate backup. Therefore it could be argued that non-digital data should be the highest priority for preservation.

At stage 3 and 4 the scope of the project was kept broad, though some themes and issues appeared to be more relevant than others and as such they were given more attention.

Previous universities that have used DAF have all ranked time management as a major issue with the methodology. Gathering survey results and conducting interviews is a lengthy process and it is advised that data collection should be timed carefully to ensure that researchers have the time available to co-operate with the project. The surveys in this project were timed to coincide with the end of the academic year when term had not yet ended but teaching workloads were reduced.

3.3 Promoting the project

Previous implementations showed the importance of encouraging survey participation. As indicated above, high participation is important in ensuring varied and representative results.

To encourage high participation:

- A £50 Amazon voucher was offered in a prize draw to survey respondents.

- A £10 Amazon voucher was offered to each researcher who took part in a follow-up interview.

- The research leaders were asked in preliminary interviews to 'buy-in' to the project and to encourage researchers in their schools to participate.

- An email was sent to all known active researchers urging participation and encouraging recipients to inform their colleagues of the survey.

- Posters were sent out to school managers for prominent display within the schools and elsewhere.

- The survey was announced on the university staff and student news web pages.

4. Findings

As mentioned earlier, data for the study were collected from two sources:

- Online Survey
- Follow-Up Interviews with Researchers

The online survey attracted 80 respondents, far more than expected. A total of 16 researchers took part in follow-up interviews. Although it is difficult to be sure how typical these participants were of the population of researchers as a whole, every school of the university was represented in both the survey and interviews (Figure 4.1, 4.2). The statistical diagrams presented in Figures 4.5 to 4.18 are based on the results of the survey only.

Figure 4.1 Survey respondents by School

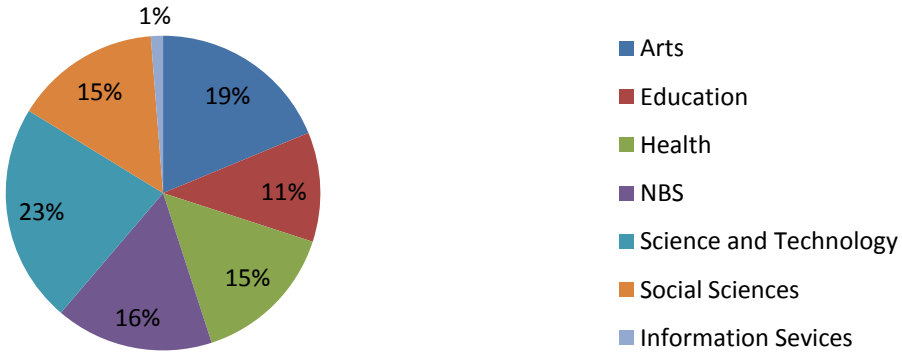
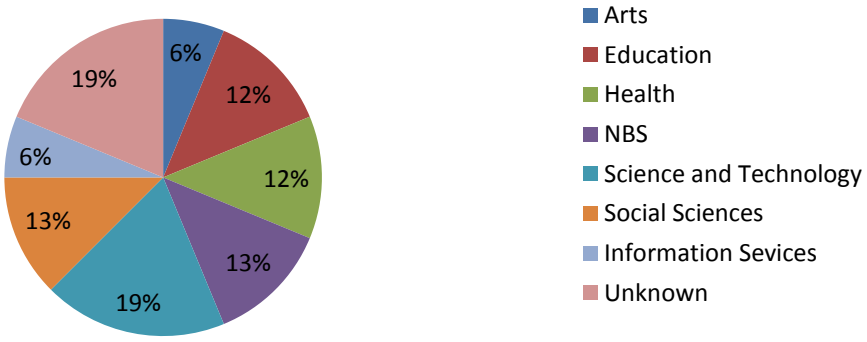
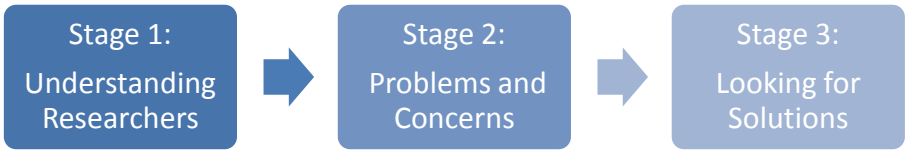


Figure 4.2 Interview participants by School



The data collected achieved three things. First, they provided a clear picture of the habits and characteristics of researchers at the university. They also brought to light the problems researchers face when working on projects. Finally, they helped to mark out possible solutions to these problems. The findings of the study will thus be presented in three stages (Figure 4.3). This provides a clear and logical order for the analysis.

Figure 4.3 Three stages of analysis



4.1 Understanding researchers and their data needs

4.1.1 Types of researcher

From the study it was found that researchers could be divided into three distinct categories, according to their research role. The way in which their research is conducted affects their needs and behaviour throughout the research life cycle.

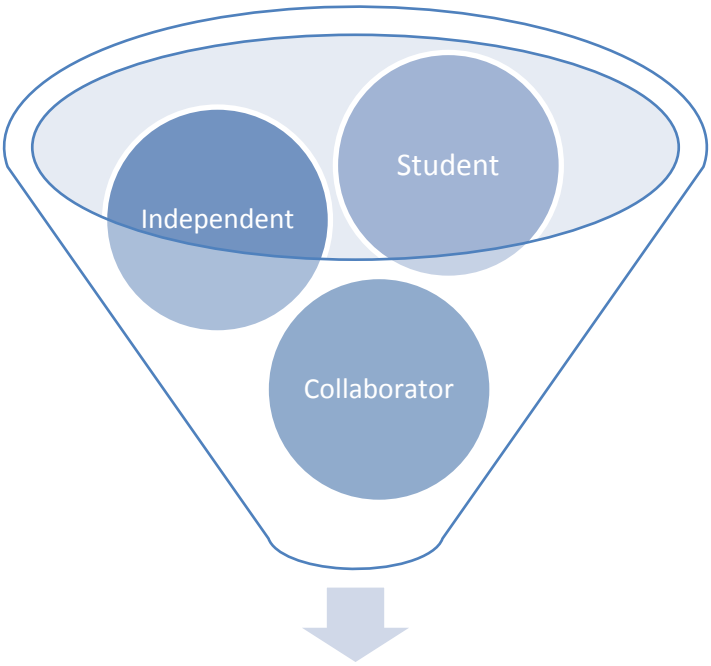
Research student:
32.5% of respondents to the survey were research students, working towards PhD or MPhil qualifications. Typically their research was carried out independently. Though not without exception, the research students interviewed were younger than other types of researcher at the university. Many of them fell into the 'Generation Y' category which is currently the subject of the major JISC/British Library sponsored project: 'Researchers of tomorrow'[12]. During the interviews, they were found to be generally less experienced in managing data than more senior researchers. For many of them it was their first time conducting research on such a large scale. Research students seemed more aware of technological developments than other researchers, and were comfortable with changes in computer software/hardware.

[12] Education for Change (2010) *Researchers of tomorrow: A three year (BL/JISC) study tracking the research behaviour of 'Generation Y' doctoral students* [online]. Available from: http://www.efc.co.uk/projects/researchers_of_tomorrow.jsp [Accessed 17th September 2010].

Independent researcher:
21.2% of respondents to the survey described themselves as independent researchers. Not working in a group, these researchers did not share data regularly. Often their notes used abbreviations that others would be unable to understand, and data files were organized in ways that would make it difficult for others to untangle. Independent researchers managed their data in ways that were tailored to their own preferences.

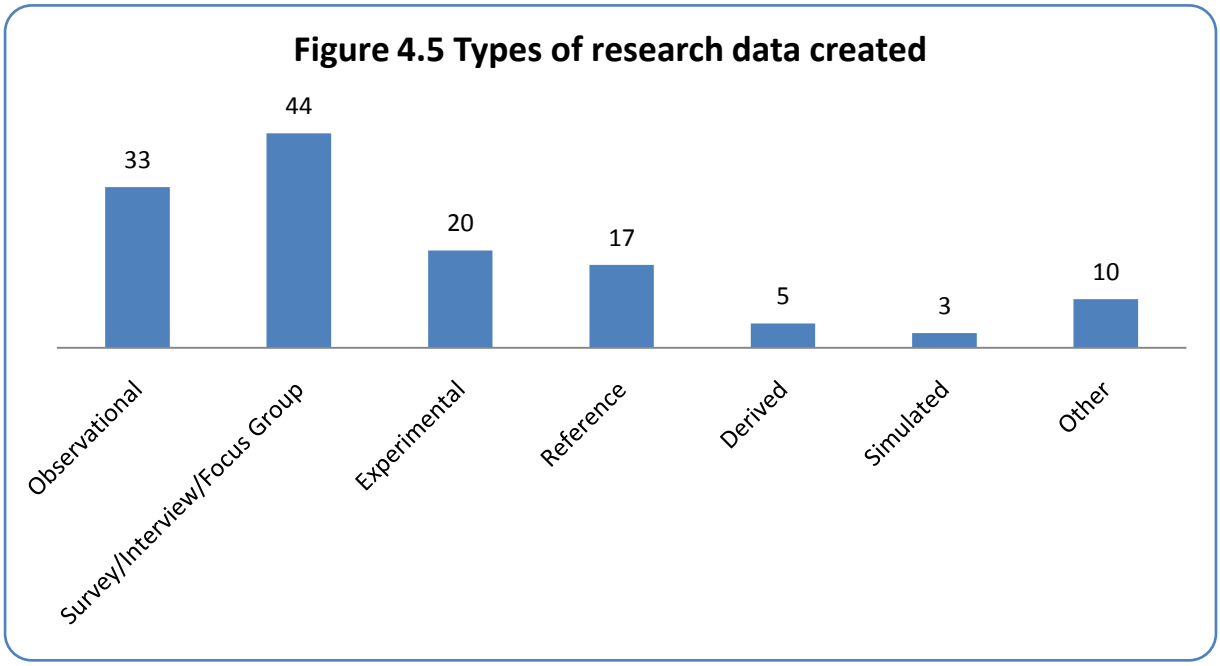
Group researcher/collaborator:
13.8% of respondents described themselves as group researchers and 18.8% as project managers. Both of these would normally work collaboratively. For them, there is a greater need to share data with their colleagues, and as such their files were likely to be organized in a way that was easier for others to read.

Figure 4.4 The make-up of the research community



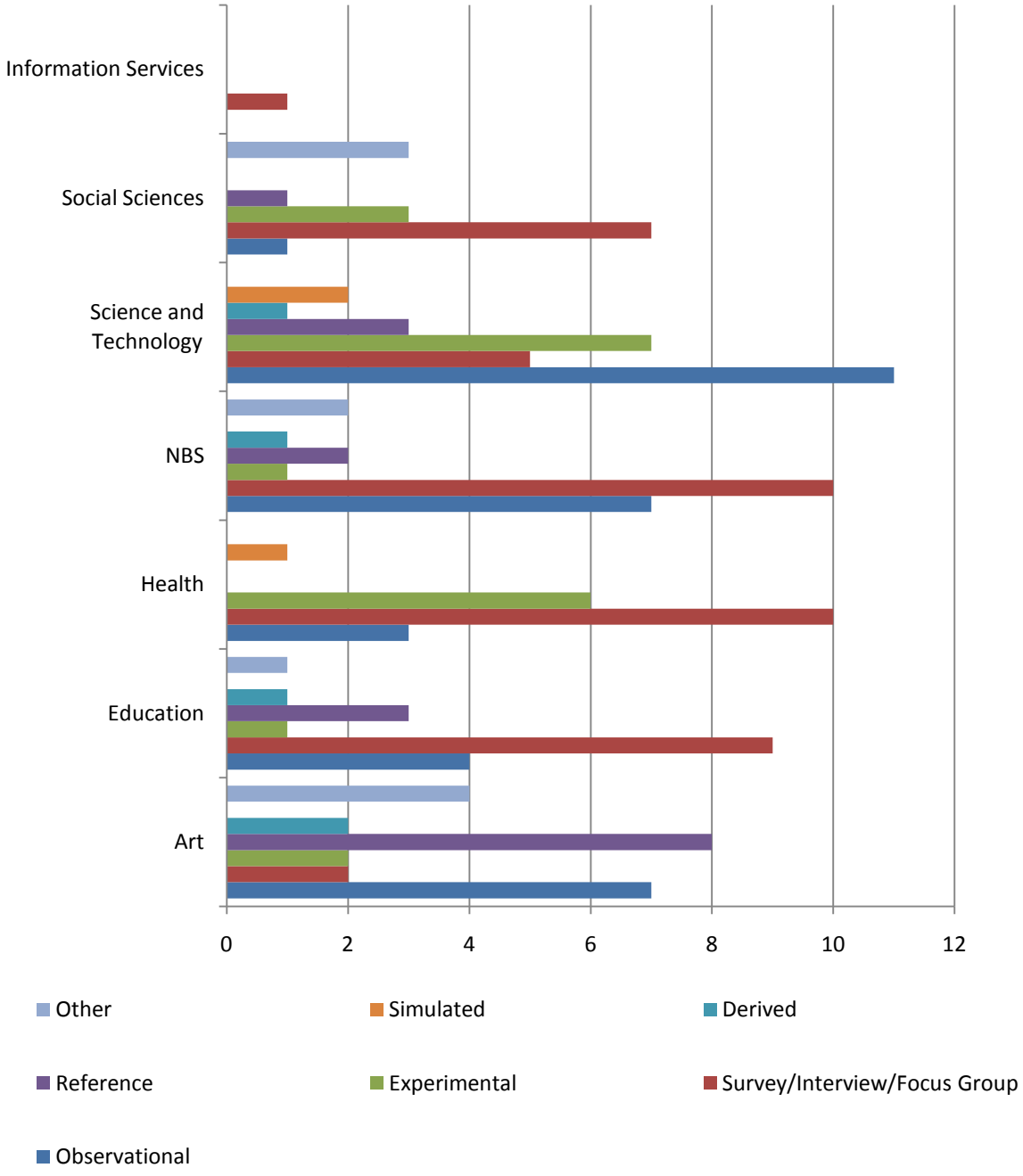
4.1.2 Types of research

A wide range of data are gathered at the university, but surveys and interviews are the most common form of data collection. Observational data, often in the form of field notes, are also frequently collected; however these data tend to play a supporting role in the research process, typically to complement results from a survey or experiment.



Understandably, the data created are dependent upon the subject of the research. Figure 4.6 makes this point clear. Notice that the School of the Arts (which contains the Division of English), is far more reliant on reference data than other schools, while the School of Science and Technology creates more experimental and observational data. Surveys and interviews are conducted for research purposes in every school.

Figure 4.6 Types of research data created (by School and Department)

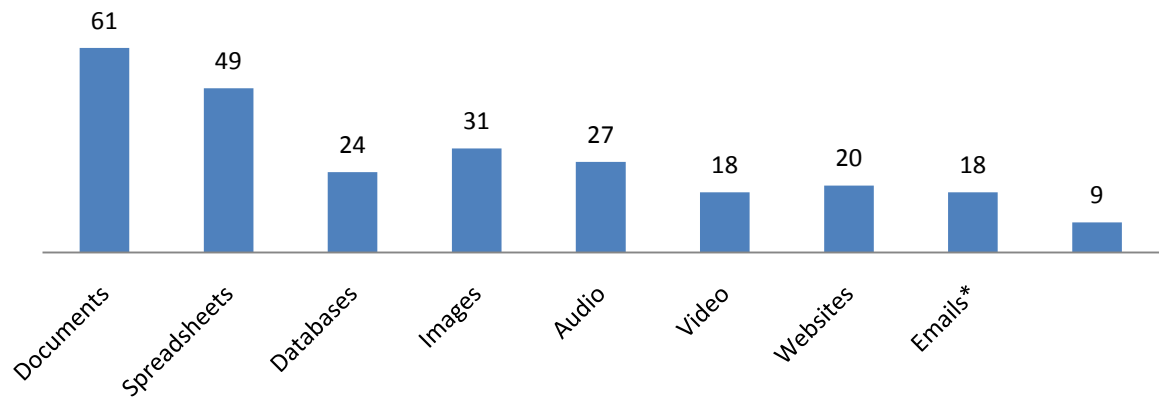


4.1.3 Data format

The format in which data are stored is dependent upon the type of data collected. Interviews are most commonly recorded on audio devices (sometimes video recordings are made) and then transcribed. Surveys creating quantitative data are typically stored in databases or spreadsheets while observational data

such as photographs and field notes are stored as images or word documents. Figure 4.7 shows the various forms of data stored by respondents to the survey.

Figure 4.7 Forms of data object



**It was unclear as to whether 'emails' referred to files stored as email attachments, or the use of actual emails to create a dataset.*

The format in which data are stored can greatly affect their accessibility. Researchers make widespread use of Microsoft software (MS Word and MS Excel) for creating text documents and spreadsheets. This is advantageous as it allows researchers to collaborate with each other easily. The similarities between Microsoft software products and the high frequency of their use enable researchers to become very familiar with product features and layout so that use of the programs becomes second nature. The format does not pose as a barrier between the researcher and the data and this allows for more effective analysis.

Figure 4.8 Primary format of documents

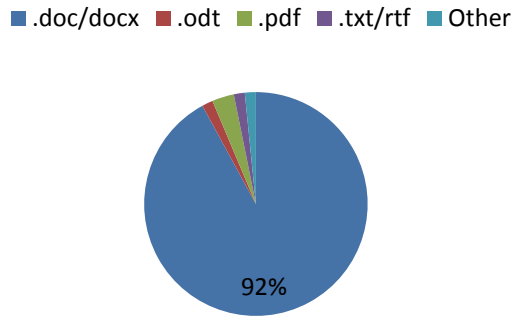
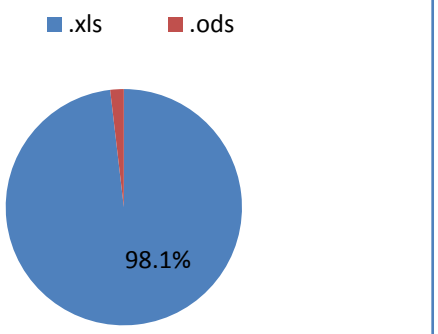


Figure 4.9 Primary format of spreadsheets



The transition from Microsoft Office 2003 to the newer 2007 version has caused some compatibility problems for researchers. Those working from home with the

older version of the software have found that the appearance of their work is sometimes altered when transferred from one version to the other. One interviewee recalled that he was unable to access his work on the older version of MS Word as he had saved his work in the wrong format (.docx as distinct from .doc). Having made this mistake once, he learned from his experience and reports that it has not been a problem since.

Some of the more senior researchers talked of a 'steep learning curve' when making the transition from MS Office 2003 to 2007, but everyone interviewed was positive about the change. It should be noted here that a newer version of the software has been released (MS 2010) and is already being used by some researchers on their personal computers. The newest version however is quite similar to MS Office 2007 and so no compatibility problems should be expected.

Unlike text documents and spreadsheets, it appears from the survey data that there is no universally adopted format for databases, although SPSS is the most widely used. In all but one of the interviews with researchers that had created databases, SPSS was the software of choice. One common complaint amongst researchers was that SPSS is updated on an annual basis, and they often have problems accessing databases that have been created on older versions. This is normally the result of researchers working from home and storing their work onto their personal hard drive instead of the university shared drive.

Figure 4.10 Primary format of databases

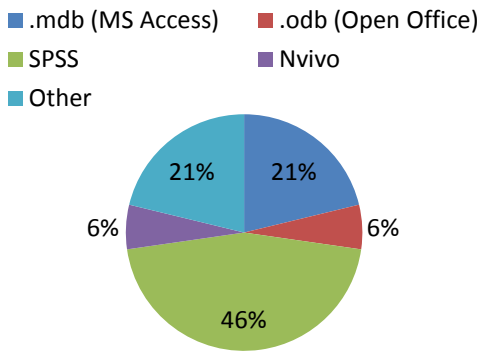
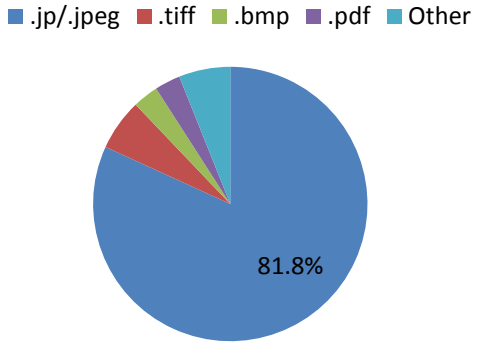


Figure 4.11 Primary format of images



With respect to the digital storing of images, it is encouraging to see that researchers at the university are predominantly using one common format (JPEG) for their data. JPEG files are the standard output from many digital cameras and can be opened using a wide range of software products (e.g. Adobe photoshop, most web browsers and most picture viewer programs), making them easier to share them with other researchers and allowing researchers to access their data from any computer, either at the university or at home. The

size of most JPEG files is not excessive, since the format is compressed. The downside of the format is that repeated editing and saving of a JPEG file will result in a loss of data and therefore of picture quality.

The storage of audio and video data appears to be a far more complex issue. There appears to be no consistency in the format chosen by researchers for either video or audio recording and this could lead to compatibility problems. With respect to audio files, .wav files are sonically superior to .mp3, but take up far more space on the computer. While .mp3 files are smaller and thus easier to transport from one device to another, they are not compatible with all CD devices. Microsoft's Windows Media Player is found on every computer in the university and plays .wma files; however these do not always convert well to CD or mp3 devices.

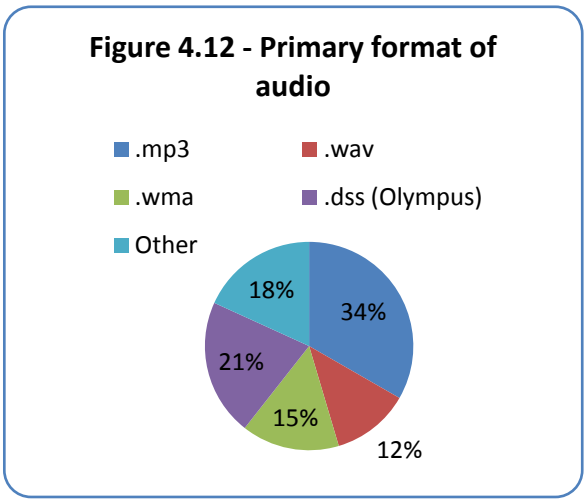


Figure 4.12 - Primary format of audio

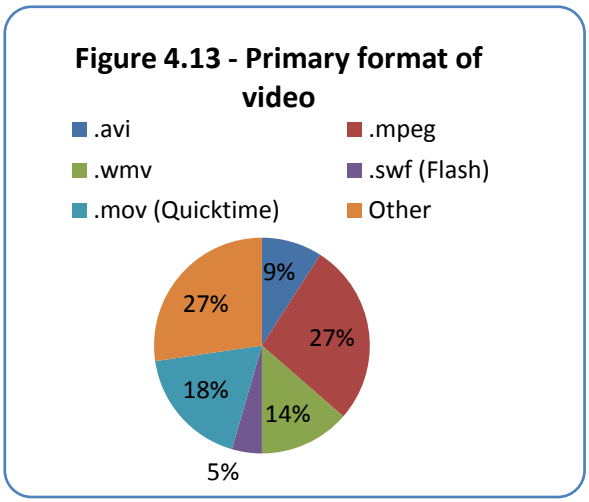


Figure 4.13 - Primary format of video

4.1.4 Data storage

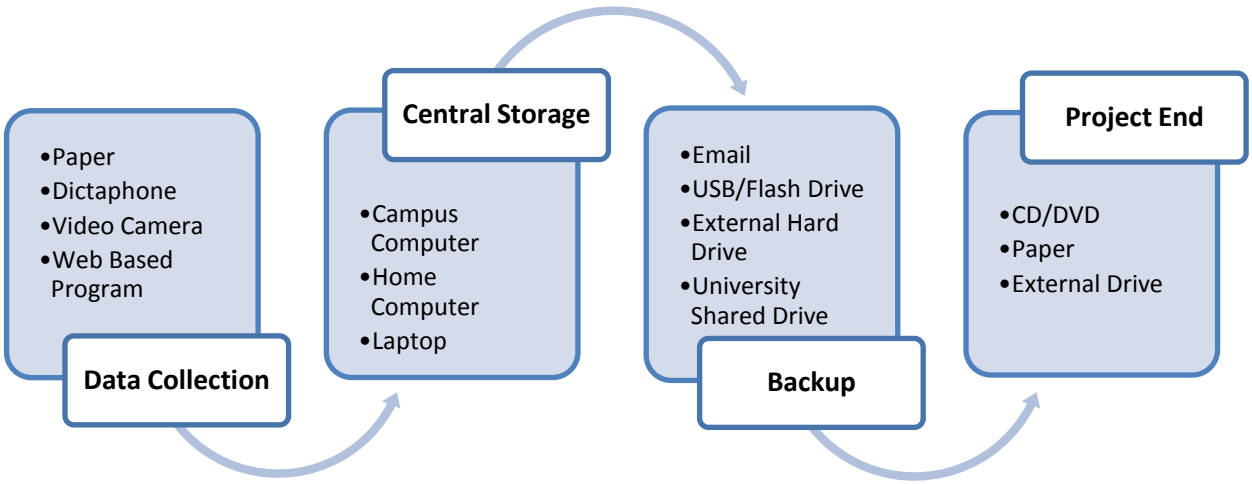
Results from question 10 of the survey, *'What are the principal media on which your research data are stored (not including backups)?'* are somewhat confusing as they contradict some of the accounts given by the interviewees. There was no limit on the number of answers respondents were allowed to give to this question, and this may explain the results. For instance, 45 respondents gave 'USB/Flash Drive' as one of their answers; more than any other device. Yet, based on the accounts of the interview participants it is hard to believe that any researcher would use a USB/Flash Drive as their *principal* medium for data storage. It is true that USB/Flash Drives are popular amongst researchers, but as a method of backing up data only. In retrospect, the design of question 10 was flawed, and this is highlighted by the response of one participant:

> *'This question is bad as the data storage method depends on the project. And basically many of the answers are applicable'*.

This experience should be noted for future related studies.

Although answers in the interviews to questions surrounding the issue of data storage were varied, a number of common themes were found. Researchers appeared to view storage devices in a hierarchical manner and to choose different storage methods at different points in the research project life cycle.

Figure. 4.14 – Data storage during the research life cycle



Data Collection:
The first requirement for data storage occurs at the data collection stage. Some researchers use paper as the initial storage medium, for example if respondents are answering a printed survey, but others gather data directly to digital media (for example to a central server from a web-based survey or to a memory card from a digital camera). Those who make recordings of face to face or telephone interviews will usually save them onto the memory of an audio or video recording device. Raw data are especially valuable to the researcher, in that they contain the fullest and most complete record, but raw data are also the most vulnerable since backup copies will not yet have been made.

It is especially important that raw data are kept securely since they will not yet have been redacted and may contain extremely sensitive information.

Central Storage:
At the data analysis stage it is usually necessary to transfer the raw data from the initial data collection device to a central storage location. This is the device on which the data object is saved first during the analysis process. Some form of conversion or translation to a new format (e.g. from survey database to Excel spreadsheet) may occur. Researchers usually keep the most up to date copy of

a data set on the central storage device. Typically, this will be either the hard drive of a campus computer, or the hard drive of a laptop.

From a data management point of view, storage of the only copy of a data file on a laptop is inherently risky unless the researcher implements a programme of regular security backups. Even then, laptops are more vulnerable to loss or theft, so ideally these should not be used as the primary storage device.

Some researchers save their work directly to the university shared drive but this practice is not common.

Backup:
This is often a portable device onto which a backup copy of the data is stored so that it can be worked on elsewhere. Many researchers prefer to use a USB/Flash Drive to store backup copies, or to save a data file as an attachment on an email account. Increasingly, external hard drives are being used to backup data. The shared university drive is also used by researchers to backup their work, but it is not accessed remotely by many researchers.

Project End:
CDs are commonly used by researchers at the end of a project to archive data that are no longer considered to be of current use to them. External hard drives fulfil the same purpose. When it is necessary to store paper-based records containing participants' personal information beyond the end of a project, locked filing cabinets are generally used. Sometimes (quite often in the case of research students) confidential information may be stored by researchers at home.

4.1.5 Data security

Over 40% of the survey respondents had at some point lost research data which had not been backed up, with roughly equal numbers blaming hardware failure, software failure and human error. That said, most (85%) researchers stated that they backed up their data regularly, mostly on a weekly, monthly or *ad hoc* basis.

Three quarters of researchers take security measures to protect their data, for example, password protecting individual files, storing them in a password protected account or encrypting the data. On university computers the necessity to log in ensures that a basic level of security is in place. Researchers' motivations for taking security measures include personal concerns (e.g. data are not ready to be released; protect own intellectual property) as well as ethical concerns (e.g. data contain personal information; funders' ethical requirements). Researchers who routinely deal with sensitive data are very attuned to the need for confidentiality and taking care of data.

4.1.6 Sharing data

Depending on the type of researcher and the nature of their research, data are shared via different methods, and with different people. The following paragraphs illustrate typical patterns of behaviour.

Research students tend to work individually, and usually do not need to share their data with anyone other than their supervisor. USB/Flash drives are often used for sharing data, especially if the file size is large. If audio and video files need to be shared, they are typically given to the supervisor on CD or DVD. Email is also popular, as it removes the need for the researcher to physically meet with the supervisor for the handing over of data. There are limits however to the size of files that may be attached to emails. (At The University of Northampton the limit for staff email attachments is set to 10MB, but this will vary from one recipient's email provider to another.)

Independent researchers too, have little need to share data on a regular basis. As such, email is a sufficient tool to use for data sharing on the occasions that they require to show their statistics to others.

Group researchers also use email regularly to share data. This can cause problems when researchers are working on and updating the same single file, as they find that they are constantly sending and receiving newer versions of the work. This highlights the need for effective version control of documents and folders. In one example, a researcher found returned to an earlier project after some time had elapsed to find several document folders on their computer with labels such as 'dataset for project', 'updated dataset for project', and 'newest dataset for project'. These labels, while meaningful at the time of creation, were subsequently unclear and confusing.

The School of Science and Technology has its own shared server where researchers working on a team project can access the same files. This is useful as it ensures that all researchers are working on the same version of the data, and it is always up to date. One researcher from the School of Science and Technology stated in his interview that this service has been invaluable to his research work, but he noted that as it is limited to his school, he is unable to share data in this way with researchers from other schools. In fact, this is incorrect, since if it is a university provided shared area, further users can be added as required.

A number of researchers collaborate with researchers from outside the university, both nationally and internationally. Email is generally used for sharing data in these cases, however as Example 1 below demonstrates, this is not without exception.

Example A

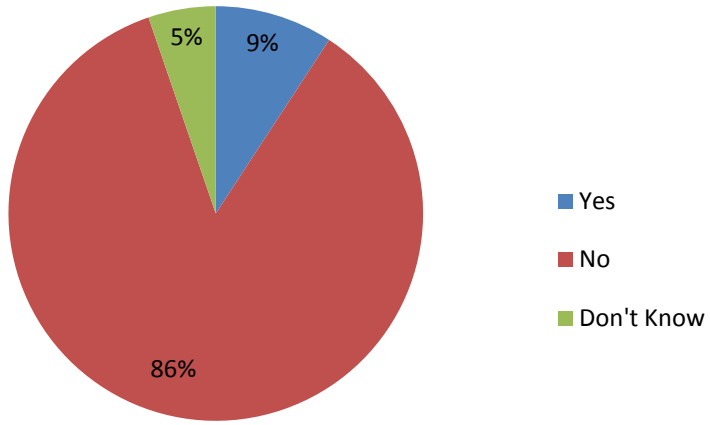
One project manager is currently heading a team of nine researchers spread across two countries. Due to the highly sensitive content of the research, and a strict confidentiality agreement with the project funders, no data can be sent over the internet. Instead, data must be exchanged by hand, requiring the project manager to travel abroad on a weekly basis.

In this situation there may well be a technical solution that would satisfy the funders, for example the use of a secure web service which is password protected. Such a service could even be linked to specific ISP addresses if required.

4.1.7 Open access to data

As mentioned at the start of this report, funding bodies are increasingly demanding that recipients of public funds should allow their research data to be openly available to the public following a project's completion. A dissemination strategy has become a normal requirement for a funded project. In question 27 of the survey respondents were asked whether or not their funders had ever requested that their research data should be made open access. The results of the question are shown below (Figure 4.15).

Figure 4.15 Have you ever applied for funding from a body that required some degree of open access to be provided for your research data?

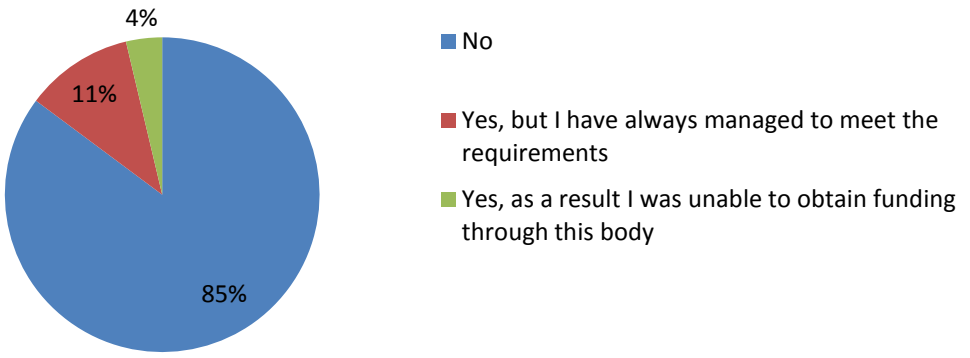


Note that when those that answered 'yes' were asked to give further details of the open access conditions, most responded with answers relating to a need for proper dissemination, or simply a need for the end results to be published in a

report. Only 4% of respondents described a genuine requirement for open access to data.

This value seems low, but may be explained by the profile of research at The University of Northampton. The majority of research at this Northampton is undertaken without the benefit of funding from the major Research Councils so it is not constrained by their mandates. Moreover, given the sensitive nature of much research in health, education and social sciences, there is less likelihood that other research funders sponsoring research in these areas will demand that data are made publicly and freely available.

Figure 4.16 Have you experienced difficulties in meeting these open access requirements?



Of the few researchers that have been asked to provide open access, even fewer have experienced problems in meeting these requirements.

Given the results above, it is unsurprising that open access to data is not a high priority for many researchers at Northampton. This could however change. The RCUK Statement of Expectation on Societal and Economic Impact[13] states that researchers should "take responsibility for the curation, management and exploitation of data for future use". Optimal exploitation of research data will be achieved only if datasets are accurately described and easily accessible. Information Services staff could play an important role in raising awareness of open research data, demonstrating the benefits of data sharing and preparing for the longer term preservation of research data.

[13] Research Councils UK (2010) *Research Councils UK mission and statement of expectation on economic and societal impact* [online]. Available from: http://www.rcuk.ac.uk/innovation/rolerc/missionsei.htm [Accessed 24th September 2010].

4.2 Problems and concerns

From the information given by respondents to the survey and participants to the interviews, seven main areas of concern have been identified. These are issues that were raised by more than one researcher, suggesting that they are common to members of the research community.

4.2.1 There is uncertainty surrounding the ownership of data

When researchers where asked the question, *"Who owns your data?"* they were often unable to give a clear answer. This was equally true of both research students and more experienced researchers. There was also very little consistency to the answers given. For instance a number of researchers felt that the data were owned by the university while others considered themselves to be the sole owners of the data. Philosophically, one researcher explained, *"the participants own the data; I am simply giving them a voice".*

Researchers appear a little clearer about copyright laws with relation to articles published in journals. One researcher said ruefully, *"you sign your rights away to get an article published".* In general, all researchers seemed to be aware of the strict limitations placed upon finished articles by publishers. Research students, many of whom were hoping to publish work for the first time, knew the least about publishing rights.

So far, confusion over the ownership of research data has not led to any serious problems at the university. This could change however, in the light of recent experiences at the University of East Anglia[14] and Queen's University Belfast[15].

---

[14] JISClegal information (2010) *Call for openness in research data* [online]. Available from: http://www.jisclegal.ac.uk/Default.aspx?tabid=243&id=1464 [Accessed 28th September 2010].
FOI decision notice available from: http://www.ico.gov.uk/upload/documents/decisionnotices/2010/fer_0238017.pdf [Accessed 28th September 2010].
For a readable summary of the 'Climategate' incident see Carrington, D. (2010) *Q&A: 'Climategate'* [online]. Guardian.co.uk. Available from: http://www.guardian.co.uk/environment/2010/jul/07/climate-emails-question-answer [Accessed 28th September 2010].

[15] JISClegal information (2010) *University must disclose raw research data* [online]. Available from: http://www.jisclegal.ac.uk/Default.aspx?tabid=243&id=1604 [Accessed 28th September 2010].
FOI Decision notice available from: http://www.ico.gov.uk/upload/documents/decisionnotices/2010/fs_50163282.pdf [Accessed 28th September 2010].

Both have been the target of Freedom of Information (FOI) requests for research data. The FOI Act will regard almost all research data to be 'held' by the university even if not necessarily owned by it. This is an area where some clarification would be useful.

4.2.2 The shared drive (R:) is underexploited

The university shared drive suffers from a poor reputation and a lack of clarity regarding its functionality. A small handful of researchers recalled in interviews that the shared drive is sometimes unavailable. Although the actual occurrences of this are minimal, they have caused the interviewees enough frustration to influence their data storage habits and cause them to refrain from using the shared drive as their principal method of storing data.

An error message that appears on the university network warning users that their disk space limit has been exceeded has also deterred researchers from using the shared drive, as they are unsure of the amount of space available to them. The amount allocated is usually flexible and can often be increased on demand. However if more space is requested, staff may be asked to consider removing duplicate and superseded copies of files as well as unused images and videos from secondary sources.

One researcher was even unaware of whether or not the shared drive is backed up.

A number of researchers are unaware that the shared drive can be accessed remotely over the internet and expressed that this would be *"a good service for the university to offer in the future".* Other researchers that are aware of remote access to the shared drive find that it is not a convenient way of accessing data.

Example B

One researcher explained that he doesn't access his shared drive from home, though this would be a useful feature for him. He knows that it is possible to access the shared drive remotely, and he even tried to do it once by following the instructions of a colleague, but he was unsuccessful. This was over a year ago and he hasn't tried since. Instead, any work he does need to take home is saved on a USB/Flash Drive.

In addition to this misinformation, there is suspicion amongst researchers over who has access to the files stored on the shared drive. On more than one occasion, interview participants, unprompted, expressed a fear of *'moles'* sifting through their folders and viewing confidential data. One researcher recalled a time when Academic IT Support *'took control'* of his cursor to solve a technical

problem he had at that time. This alarmed the researcher and made him think twice about the shared drive and who has access to it. It should be noted that the practice of remotely controlling a computer occurs only in exceptional circumstances and with the permission of the user. Moreover, Academic IT Support cannot access any file which is password protected or encrypted.

4.2.3 Data are neglected once a project is complete

It was found in the interviews that most researchers were keen to keep the data following the completion of a project. Unless requested otherwise by funders, researchers seem intent on keeping their data indefinitely. This is not realised however for a number of reasons.

First, the majority of researchers have no system in place for archiving their work upon the completion of a project. During the research process, data can become scattered across various storage devices and several versions of the same data file might exist. This can cause problems for researchers that wish to return to these datasets at a later date. During the interviews, a number of researchers described difficulties that they faced in locating the correct and most up to date version of a file when returning to work months or sometimes years later. Confusing or unclear labelling was often cited as the culprit for this.

Audio and video files are often too large for researchers to store on their computer hard drives once a project is complete. Instead, it was found that many researchers will transfer the files onto CDs, DVDs, and USB/Flash Drives. This can cause problems as these media are all prone to degradation and can easily be lost.

Researchers who choose to store their data at home rather than use the university shared drive have sometimes found that their files can become incompatible with the university computers due to software updates. The example most regularly given by researchers was the database program SPSS, which one interviewee claimed, "is updated every year".

4.2.4 Data management is guided by intuition

It was found during the interview process that a huge variety of data storage and management methods are currently being used at the university. This is understandable, as the range of research topics is equally vast. However, some methods are more effective than others. Some PhD students did say that they had received advice on storing data from their supervisors, however most interviewees seem to "go with what feels right". Most researchers appear to be satisfied with their data management practices. Data management is not regularly discussed by researchers, and so they have nothing with which to

compare their own performance. In the context of data management, researchers' behaviour is satisficing rather than optimising.

Example C

The extent to which data storage and management methods can differ between researchers was made clear during interviews with two researchers that share the same office. It was found that while one researcher makes regular use of the university shared drive, the other has never used it. Both researchers were unaware of the other's data management practices. They said that it is not something they have ever discussed before.

4.2.5 Researchers are unaware of the services on offer

All researchers expressed satisfaction with the support provided by Information Services. One researcher stated that, "support is on tap... if there is anything you need, you just ask". Whilst it is encouraging to hear such positive feedback, IS should not be complacent. Indeed, there are two main causes for concern. First, it is likely that researchers who are not in the habit of asking for assistance may be missing out on services that could be beneficial to them. Second, as indicated above, some researchers are uncertain of the services they are entitled to.

4.2.6 Data are being collected in out-dated formats

It was found during the interviews that some data are still being stored on floppy disk. This is worrying since the university is currently updating its staff computers and the newer systems no longer accept floppy disks. For most researchers this is not a problem, they will simply migrate transfer the data from these disks onto newer formats. In some rare cases though, the use of floppy disks is unavoidable.

Example D

One interviewee from the School of Health showed concern for a piece of equipment currently used for experiments, which records its results directly onto floppy disk. He explained that it is becoming increasingly difficult to find computers in the School that accept floppy disks. To replace the machine would apparently cost thousands of pounds.

A number of researchers experience problems in storing paper due to a lack of space. It was found in the interviews that researchers regularly print hard copies of documents, as they find it is easier to read work and make annotations in this way. For many researchers, filing cabinets are constantly at maximum capacity

and documents are frequently being destroyed to make room for new ones. In most cases, the documents destroyed are not vitally important to the researcher, but all researchers interviewed stated that they would prefer to keep these documents for longer.

Good records management practice suggests that paper copies of documents should be created for immediate use but then disposed of as the research progresses. Not only would this save space in physical storage, but also it is possible to annotate or 'track changes' on electronic documents, and new and up to date versions can be easily retrieved as needed.

It was found that some interviews are still being recorded onto tape using analogue Dictaphones. This is not ideal as tape can only be backed up in real time, a lengthy process that time constrained researchers will usually skip. This means that the raw audio data is not backed up. Also, tapes are prone to degradation over time, making the raw audio data extremely vulnerable.

Information Services could have a useful role in advising and facilitating migration of data from one format to another.

4.2.7 Transcribing interviews is a problem for researchers

The transcription of interviews is a hugely time consuming process for researchers. A number of researchers have used professional transcription services in the past; however for many of them the results have been disappointing. It was explained that, unless the transcriber is familiar with the subject matter and the terminology used, the interview content can be misinterpreted and its meaning may be lost entirely. Additionally, research students, often conducting interviews for the very first time, appear unsure of where to find transcription services.

Surprisingly, no comment was made about the potential infringement of Data Protection legislation if the researcher uses a third party to transcribe an interview. If an external transcription service is to be used, it is important that the interviewee is informed of this in advance of the interview and that the transcription service is given clear instructions regarding data handling and security. Failure to do these could mean that the researcher is acting illegally.

4.3 Looking for solutions

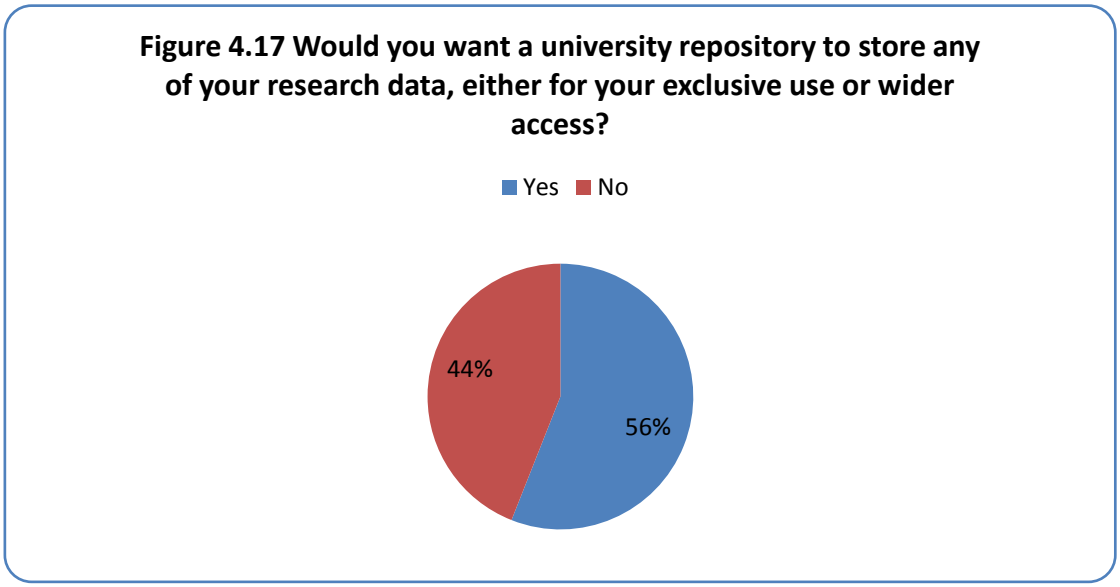
4.3.1 Data repository

It was noted in the rationale for the project (see Section 1.1) that the university's research repository, NECTAR, was technically capable of storing data sets on behalf of the research community. This could be advantageous in

enabling researchers to store their research data safely and securely in a space that is backed up regularly and in freeing up space on personal computer hard drives. It could also provide researchers that are required to offer open access to their data, a means by which to do so, while collaborating researchers could use the repository to share files more easily with colleagues.

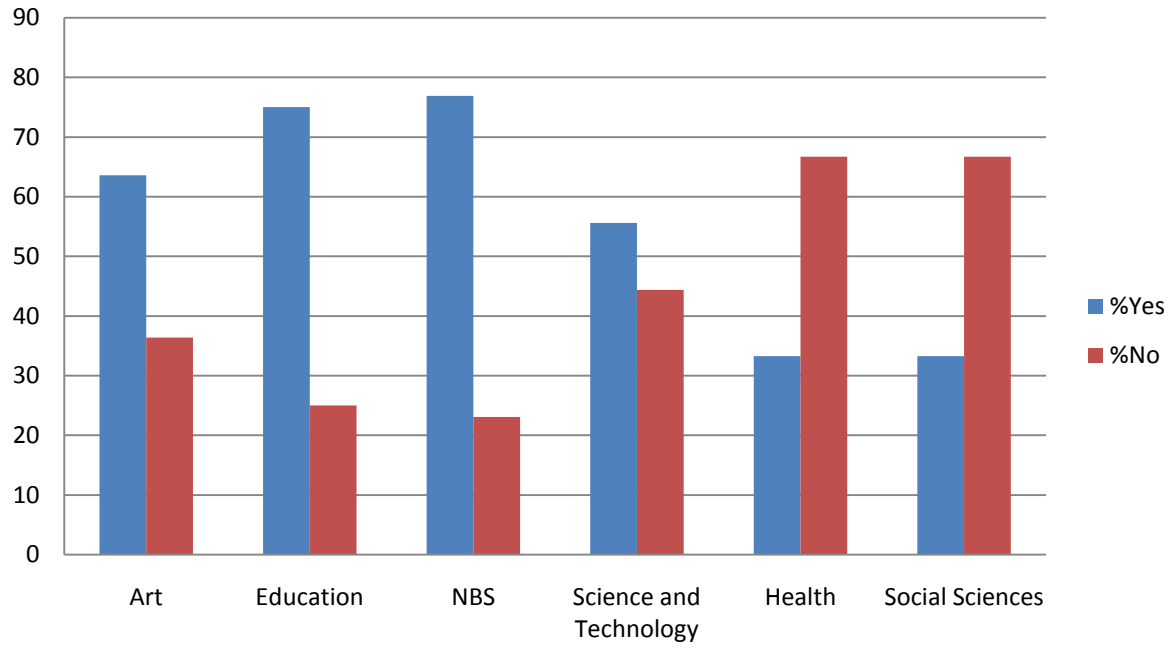
The idea of a repository however does not appear to be hugely popular amongst the research community. As Figure 4.17 shows, only 56% of survey respondents stated that they would want such a service. Of those that said 'yes', relatively few wanted *all* of their data to be stored in the repository but a significant minority were interested in having at least some of their data stored there and retained until and beyond the end of the project. Only seven researchers were interested in having any of their data stored in perpetuity.

Figure 4.17 Would you want a university repository to store any of your research data, either for your exclusive use or wider access?



Interestingly, opinions differ on this between the schools. As Figure 4.18 shows, the most opposition to the repository comes from researchers in the schools of Health and Social Sciences, where the majority of research is in the form of surveys and interviews, and the content is often of a sensitive nature.

Figure 4.18 Would you want a university repository to store any of your research data, either for your exclusive or wider access? (by School)



The main concern of researchers with regard to the proposed repository was the thought that others would have unrestricted access to their data. The majority of researchers were explicitly against this idea. They did however express more willingness to be contacted personally by other researchers wanting access to their data. With this in mind, it may be more feasible to create a searchable index of metadata describing the research data.

Researchers see the potential benefits of having a means by which to share files more easily, but for many (particularly research students and researchers working alone) it is doubtful that they would have the need to use such a service. Also, researchers appear more interested in using the repository as a means for storing data from completed projects as distinct from works in progress. As one researcher explained, "*it would be like a savings account rather than a current account*".

4.3.2 Training

During the interviews it was found that the vast majority of researchers have not received, or at least do not recall having received, any training relating to data management. Moreover, there are currently no guidelines in place that explicitly address research data management.

Information Services has offered generic records management training and advice for the past three years. Although this already encompasses the

management of data, this perhaps has not been recognised by the research community. A variation of the existing offering, with a focus on research data, might attract a larger audience of researchers.

The core records management training might usefully be supplemented by a range of formal and informal training and support activities. These would go some way towards mitigating the problems identified above. Some possible aspects of the training programme are outlined below.

4.3.3 Other possible solutions

During the interviews researchers were invited to suggest ways in which they could be supported in managing their data. The following suggestions were made.

i. Researchers should be provided with their own web space. This would allow them to store their work online, and display it publicly. One researcher noted how useful this service was to her when it was provided at a previous university.

ii. The university should appoint a 'statistics officer' to support researchers without previous experience in quantitative analysis. Lack of confidence

in statistical analysis was cited as a reason for reluctance to offer open access to datasets.

iii. A list of recommended transcribers could be available to those seeking professional transcription services.

5. Lessons learned

Having benefitted from the experiences of other DAF projects, we were able to initiate, plan and conduct this project over a very short timescale (eight weeks from start to finish – see Appendix 1). On the whole we were more than satisfied with the result but, as with most research projects, in hindsight we might have done some things differently. These are some of our reflections on the process.

- Early interviews with research leaders aimed principally at encouraging 'buy-in' proved extremely useful for later participation rates. If our project had requested these interviews during stage 1, we would have had more research leader input in time for the pilot survey.

- Because of the very short lead time for the project (three days) we were unable to go through the formal university ethics approval process, although we did seek (and follow) advice from the Chair of the university's Research Ethics Committee. As researchers themselves, participants were fully aware of the ethical issues surrounding this type of project; we believe we conducted the research in an ethical manner but full ethical approval would have been preferable.

- We sought permission from participants for the use of their anonymised responses in publications resulting from the project. In doing so we did not differentiate between publications for internal and external consumption. In retrospect, this might have been appropriate.

- Bristol Online Surveys is an easy-to-use, if somewhat limited, system for survey publishing. It is certainly preferred to QuestionMark Perception, given Southampton's experience.

- We were fortunate that although our survey was longer than most other DAF implementations, the response rate was fairly high. Some of the more detailed technical information has not been included in this report but will inform future service developments. We are grateful to members of the research community at Northampton who cooperated so willingly.

6. Recommendations

The Research Data Survey identified both good and less good practice in the management of research data. It was apparent that researchers held a number of misconceptions with respect to the services available to them for storing and curating their data and in many cases were unaware of the training opportunities already on offer. In the light of this, the following recommendations are made.

Recommendation 1. This report to be presented to senior research managers and to the university's Research Committee for discussion.

Recommendation 2. A university research data policy to be drafted and approved by Research Committee. This policy should be guided by the criteria set out in the Data Seal of Approval, following the precedent of the University of Bath[16].

Recommendation 3. With appropriate reference to both contractual arrangements and copyright law, the university to clarify its position on the ownership of research data and other research outputs generated by staff and research students at Northampton. This information to be disseminated to all researchers.

Recommendation 4. Information Services, in conjunction with the Graduate School, to develop and promote training sessions on 'Data management for researchers'. Based on existing records management training, these will focus on the specific needs of researchers. A version of this training session might usefully be incorporated within the research students' mandatory induction week.

Recommendation 5. Information Services to produce a guide to research data management. Contextualised by the research life cycle, this guide will summarise the researcher's legal obligations, describe good practice in research data management and clarify commonly held misconceptions.

Recommendation 6. Information Services to further develop and disseminate expertise in preservation planning (including analysis of risk, identification of requirements and creation of preservation strategies and action plans) to support

[16] Ball, A. (2010) *Review of the state of the art of the digital curation of research data* [online]. Bath: University of Bath. Available from: http://opus.bath.ac.uk/18774/ [Accessed 17th September 2010].

researchers wishing to store and access their data over the medium to long term.

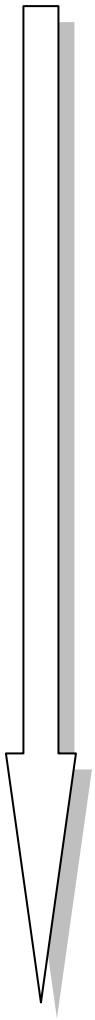
Recommendation 7. The Research Support Specialist to present the findings of the report to Schools and Research Centres, together with advice and guidance in line with the new research data policy and information concerning the ownership and exercise of rights to research data.

Recommendation 8. This report to be deposited in NECTAR.

Recommendation 9. The Data Curation Centre to be informed of the project and invited to create a link to NECTAR from the DAF website so that other institutions may read about Northampton's implementation of the framework.

Appendix 1 – Project Timeline

10/05/2010	Project launched
14/05/2010	First meeting of Project Board
21/05/2010	First meeting with research leader
	Pilot survey launched
24/05/2010 to 28/05/2010	Meetings with research leaders and Information Systems developers
28/05/2010	Live survey launched
4/06/2010	Second meeting of Project Board
7/06/2010	Handover from Sam McKenney to Edward Alexogiannopoulos
9/06/2010 to 10/06/2010	Pilot interviews with researchers
14/06/2010 to 24/06/2010	Interviews with researchers
21/06/2010	'Live' survey closed
21/06/2010 to 02/07/2010	Data analysis and writing up
30/06/2010	Third meeting of Project Board
02/07/2010	Project finished



Appendix 2 – Survey Questionnaire

The 'pilot' and 'live' surveys were composed and distributed using Bristol Online Surveys (BOS). BOS is a highly intuitive tool to use but does not allow certain features, such as branching, within the survey. This imposes a significant restriction on the flow of questions.
A copy of the 'live' survey is shown below.

Page 1:

Welcome to the Research Data Management Survey

This survey is for research-active staff and research students at The University of Northampton. It is designed to build a better understanding of the data held by researchers in The University of Northampton, of researchers' current data management practices and of their needs.

The results of this survey will inform a data management policy for the university and procedures to support the data management needs of the research community (for example to provide security, access, and long term storage of research data). This in turn will enable members of the university to satisfy the stricter access requirements to research data now specified by many funding bodies.

It would help us greatly if you respond to this questionnaire even if you do not currently hold any research data. None of the questions are mandatory; please skip any questions that are not relevant to you.

This survey is a maximum of 32 questions and should take no more than 15 minutes to complete. It can also be saved at the end of any page and continued later. Please note however that you will not be able to change your responses on previous pages.

Once you have completed the survey, please indicate whether you wish to be entered into a prize draw for £50 of Amazon vouchers. We will also be giving £10 of Amazon vouchers to everyone who participates in a short follow-up interview.

Thank you for your time.

Page 2:

Data Protection

Any personal data collected during this survey will be retained only as long as is necessary to compile anonymised statistical data. After this process is complete all personal data will be destroyed in a secure manner. Cookies (personal data stored by your web browser) are not used in this survey.

It would assist the project if you were to provide your name at the end of the survey, but, if you do not wish to, please complete the survey anyway. Anonymous responses will still give supporting evidence for the rest of the project.

Page 3:

Personal information

1. What best describes your main research role?
   - □ Principal Investigator/Project Manager
   - □ Member of Research Team/Group
   - □ Independent Researcher
   - □ Research Assistant
   - □ Research Support/Non-academic Staff
   - □ Research Student (PhD or MPhil)
   - □ Other (please specify):_____

2. Research group or research active area: _____

3. School:
   - □ Arts
   - □ Education
   - □ Health
   - □ Northampton Business School
   - □ Science & Technology
   - □ Social Sciences
   - □ Other (please specify):_____

Page 4:

Details of your research data

For the purpose of this study your 'research data' are data that have been collected and/or used in the course of your research at The University of Northampton. Research data can be primary data collected by you or your research group or secondary data provided by a third party. They may be quantitative or qualitative e.g. survey results, interview transcripts, databases compiled from documentary sources, images or audiovisual files.

'Research data' do NOT include publications, articles, lectures or presentations.

Data that you 'hold' describes any the research data that you store anywhere. For example: on a computer, on CDs or on paper.

4. Do you currently hold or have you ever held any research data? (If no, please skip to 'Conclusion' on page 8)
   - □ Yes, I currently hold research data
   - □ Yes, I have held research data in the past
   - □ No

Research data you 'own' describes data to which you, at least in effect, hold some intellectual property rights. Unless you and the University, or your funder, have agreed otherwise, and your research is non-commercial, you own these rights.

Even if you do not currently hold any data, we would appreciate it if you complete the remainder of the survey to the best of your recollection of data you have previously held.

5. Who owns the research data you hold?
   - □ I own all of the data I hold
   - □ I own some of the data I hold
   - □ I own none of the data I hold
   - □ Don't know

6. Do you share ownership of any of your research data with others? (select all that apply)

   - □ No
   - □ Yes, with other academics/researchers
   - □ Yes, with journals/publishers
   - □ Yes, with funding bodies
   - □ Other (please specify): _____

7. Are you currently receiving funding for a research project?
   - □ Yes
   - □ No

If yes, who are you receiving funding from? _____

8. Which of the following categories best describes the research data created in your field of research? (select all that apply)

- □ **Observational** e.g. video or audio recordings of performances or other primary sources; photographs of artistic works, historical documents etc. (researcher has a passive role)
- □ **Survey/Interview/Focus Group** e.g. quantitative or qualitative responses to survey or interview questions; oral history accounts (researcher has an active role)
- □ **Experimental** e.g. spectrometry results
- □ **Reference** e.g data cataloguing/describing other datasets
- □ **Derived** e.g data from interrelating survey data
- □ **Simulated** e.g data from a engineering model
- □ Other (please specify): _____

9. What types of research data do you hold (e.g. laboratory notes, image collections, transcripts etc.)? _____

10. What are the principal media on which your research data are stored (not including backups)? (select all that apply)

- □ Hard disk drive of computer on campus
- □ Hard disk drive of computer off campus
- □ Hard disk drive of laptop/netbook
- □ Hard disk drive of instrument/sensor which generates data
- □ External hard drive
- □ Shared drive/server (e.g. University server)
- □ Third party (including commercial data storage)
- □ Web-based service (e.g. Google Docs, Flickr, Box.net, Dropbox, Pando etc. (please specify under 'Other')
- □ CD/DVD
- □ USB/Flash drive
- □ Email client/server
- □ Floppy Disk
- □ VHS/Video Cassette
- □ Cassette Tape (Audio)
- □ Photograph
- □ Slides
- □ Microfiche
- □ On paper
- □ Other (please specify): _____

11. What formats/software do you use for your electronic research data? (select all that apply)

- □ Documents
- □ Spreadsheets

- □ Databases
- □ Images
- □ Audio
- □ Video
- □ Websites
- □ Emails (not including other formats attached to emails)
- □ Unique program/simulation written specifically for project
- □ Other (please specify): _____

a. If you store data in documents, please select the primary format you use:
   - □ MS Word .doc/.docx
   - □ OpenOffice Writer .odt
   - □ Adobe .pdf
   - □ Appleworks .cwk
   - □ .xml
   - □ .txt/.rtf
   - □ Other: _____

b. If you store data in spreadsheets, please select the primary format you use:
   - □ MS Excel .xls
   - □ OpenOffice Calc .ods
   - □ Appleworks .cwk
   - □ Other: _____

c. If you store data in databases, please select the primary program you use:
   - □ MS Access .mdb
   - □ OpenOffice .odb
   - □ SPSS
   - □ Oracle
   - □ MySQL
   - □ NVivo
   - □ Other: _____

d. If you store data as images, please select the primary format you use:
   - □ .jpg/.jpeg
   - □ .gif
   - □ .tiff
   - □ .bmp
   - □ Adobe .pdf
   - □ Adobe .ai
   - □ .svg
   - □ Other: _____

e. If you store data as audio, please select the primary format you use:
   - □ .mp3
   - □ .wav
   - □ .wma

- □ Olympus dictaphones .dss
- □ Other: _____

f. If you store data as video, please select the primary format you use:

- □ .avi
- □ .mpeg
- □ .wmv
- □ Flash .swf
- □ Quicktime .mov
- □ Other: _____

g. If you have selected 'Other' for any of the questions a-f please give details of the software or formats you use: _____

12. On average, how frequently do you update your research data during the project they relate to?

- □ Never
- □ Daily
- □ Weekly
- □ Monthly
- □ Annually
- □ Ad hoc
- □ Other (please specify): _____

13. Please estimate the total combined size of all your electronic research data: _____

14. Have you ever experienced any problems storing your research data due to the size of the files?

- □ Yes
- □ No

If yes, please give details: _____

Page 5:

Research data storage and security

15. Do you currently have any data management plans for your research data (for example, data preservation policy, data security policy, record management policy, data disposal strategy)?

- □ Yes
- □ No
- □ Don't know

If yes, please give details: _____

16. Who, if anyone, is currently responsible for managing your research data? (select all that apply)

- □ Yourself (select other options only if they are not you)
- □ Principal Investigator/Project Manager
- □ Research Assistant
- □ Research Technician
- □ Research Support/Non-academic staff
- □ Research Student (PhD or MPhil)
- □ External project partners
- □ School IT technician
- □ Information Services (at UoN)
- □ Local data centre
- □ National data centre / data archive
- □ International data centre / data archive
- □ Nobody
- □ Don't know
- □ Other (please specify): _____

If you use any external data centre or archive, please give details: _____

17. Have you ever lost research data which was not backed up? (select all that apply)

- □ No
- □ Yes, through hardware failure
- □ Yes, through software failure
- □ Yes, through human error or loss

18. How often do you take security measures to protect your research data (e.g. password protect file, store only in a password protected account/profile, encrypt etc.)?

- □ Never
- □ Sometimes
- □ Often
- □ Always

If yes, what motivates you to take security measures to protect your research data? (select all that apply)

- □ Data have commercial value
- □ Funder requirements
- □ Data are not ready to be released/concern unpublished work
- □ Protect own ideas or intellectual property
- □ Data contain personal information/have not been anonymised
- □ Ethics requirements of university/funder
- □ Other (please specify): _____

19. Have you ever been concerned that any of your research data may have been unintentionally released?

- □ Yes
- □ No

If, yes, how (e.g. emailed data to the wrong person, laptop stolen)? _____

20. Are any of your research data backed up regularly?
    - □ Yes
    - □ No
    - □ Don't know

    a. What research data tend to be backed up?
        - □ Everything
        - □ Data critical to project
        - □ Data required for publication
        - □ Don't know
        - □ Other (please specify): _____

    b. How frequently are they backed up?
        - □ Daily
        - □ Weekly
        - □ Monthly
        - □ Annually
        - □ Ad hoc
        - □ Don't know
        - □ Other (please specify): _____

    c. Where are they backed up? (select all that apply)
        - □ Hard disk drive of computer on campus
        - □ Hard disk drive of computer off campus
        - □ Hard disk drive of laptop/netbook
        - □ Hard disk drive of instrument/sensor which generates data
        - □ External hard drive
        - □ Shared drive/server (e.g. University server)
        - □ Third party (including commercial data storage)
        - □ Web-based service (e.g. Google Docs, Flickr, Box.net, Dropbox, Pando etc. (please specify under 'Other')
        - □ CD/DVD
        - □ USB/Flash drive
        - □ Email client/server
        - □ Floppy Disk
        - □ VHS/Video Cassette
        - □ Cassette Tape (Audio)
        - □ Photograph
        - □ Slides
        - □ Microfiche
        - □ On paper
        - □ Don't know
        - □ Other (please specify): _____

21. If the service was offered, would you want a University of Northampton repository to store any of your research data, either for your exclusive use or for wider access?

The hypothetical repository would offer to store whatever research data researchers volunteer (and possess the appropriate rights to volunteer) with a retention period of their choosing. The files would be stored securely with accessibility limited by default to only the researcher in question. The researcher would have the option of widening access anywhere from specific other users to full public open access.
The repository would, therefore, provide separate, voluntary facilities for: long-term storage, backups, sharing of data for collaboration purposes with colleagues, and open access. The repository would offer facilities aimed at meeting stricter requirements now made by many funding bodies.

- □ Yes
- □ No

22. If yes, how long would you want the repository to retain any of your research data, including data only accessible by you?

	None of my data	Some of my data	Much of my data	All of my data
Not at all				
Until the end of the project				
For a finite period after end of project				
Until I leave the university				
In perpetuity				

Page 6:

Research data access

23. How do you currently share research data with colleagues? (select all that apply)
- □ I never share data with colleagues
- □ E-mail
- □ Shared computer
- □ Shared drive/server (e.g. University server)
- □ Using portable storage (e.g. CDs, DVDs, external hard drive, memory sticks etc.)
- □ Web-based service (e.g. Google Docs, Flickr, Box.net, Dropbox, Pando etc. (please specify under 'Other')
- □ On paper
- □ Other (please specify): _____

24. Have you encountered any problems sharing data with colleagues? (select all that apply)
- □ No

- □ Finding suitable shared storage space
- □ Lack of file naming conventions made it difficult to identify files
- □ Lack of version control caused confusion
- □ Legal issues arising from international transfer of data
- □ Problems establishing ownership of data
- □ Time consuming to keep all colleagues constantly up to date
- □ Other (please specify): _____

25. Apart from yourself, who would you want to be allowed access to your research data?

	None of my data	Some of my data	Much of my data	All of my data
My colleagues				
My school				
The whole university				
Specified academic communities beyond the university				
Anyone (including general public)				

26. What factors would prevent your research data from being made open access to the general public? (select all that apply)

- □ None
- □ I do not believe the public would have any use for some of my data
- □ I do not have the ownership rights to share all of my data
- □ Data have commercial value
- □ Funder restrictions
- □ Data are not ready to be released/concern unpublished work
- □ Protect own ideas or intellectual property
- □ Data contain personal information/have not been anonymised
- □ Ethics requirements of university/funder
- □ Other (please specify): _____

27. Have you ever applied for funding from a body that required some degree of open access to be provided for your research data?

- □ Yes
- □ No
- □ Don't know

a. If yes, please state funder and give details: _____

b. Have you ever experienced difficulties in meeting these requirements?
   - □ No
   - □ Yes, but I have always been able to meet the requirements
   - □ Yes, as a result I was unable to obtain funding through this body

28. Do you have any specific concerns over the current management of your data or services you would like to see offered by the university to guarantee access to this data in the future? _____

Page 7:

Conclusion

29. Do you have any questions, comments or suggestions concerning this survey or data management within the university? _____

30. The second part of this project will involve exploring individual researchers' data management practices in a one to one conversation. This will help us further to establish and support your needs. A £10 Amazon voucher will be given to anyone we interview.

Please confirm if you would be willing to participate in a short follow-up interview.

- □ Yes
- □ No

31. If you are willing to be contacted for interview or wish to enter the prize draw, please tell us your name and university contact details: _____

32. Please indicate below if you wish to be entered into our prize draw for £50 of Amazon vouchers.

- □ Yes
- □ No

Page 8:

Research Data Management Survey

Thank you for completing this survey, your contribution is very much appreciated.

If you have any questions about this survey, or would like to discuss the management of research data at The University of Northampton, please contact Miggie Pickton in Information Services.

Appendix 3 – Follow-up Interview Plan

Follow-up Interview Plan

Take a copy of participants' survey responses and any other relevant information about them (e.g. generic information offered by research leaders). During the interview make certain of which information may be cited in later publication (this is to avoid ethics/anonymity problems later on) and assure those with particularly sensitive issues (e.g. unintentional data release) that no identifying information about the incidents will be reported.

Introduce project

Introduce self and then the project (an implementation of the tried and tested DAF methodology).

The goals of the interview are to:

- Investigate current data management practices and researchers' data service needs.
- Determine which services and policies would satisfy these needs, satisfy the demands of funders, and help to ensure a continuing high standard of research at UoN.

Ask if the participant is willing to have the interview recorded.

Questions

Proceed with follow-up questions to the survey as needed to clarify respondents' answers, with particular focus on:

- □ The nature of the research
- □ How research data is stored/backed up
- □ How much data is hold and in which formats/media
- □ How security is handled
- □ How data are shared/collaboration is facilitated
- □ Who the funds the research and what are their requirements (check these subsequently against JULIET)

Investigate a specific data object

Ask the participant about a particular research project and related data objects. Find out which access/funding/anonymity issues surround the objects and which services would be helpful for the objects. Complete the metadata form (see Appendix 4) for one data object.

Explore

Seek any additional relevant information the respondent wishes to volunteer and explore their ideas for future services/policies. Would other researchers in the same school share their views?

Thanks

Thank the participant for their co-operation and give them the £10 Amazon Voucher (signature for receipt required). Agree a date by which they will receive a response to any issues that have been raised which could not be answered during the interview and notify them of when they can expect to see any data management changes implemented.

Appendix 4 - Research Data Management Project Metadata Form

Name of Interviewee: **Date of Interview:**

Data Object

Title: Official name of the object, with additional or alternative titles or acronyms if they exist

Description: A brief description of the information contained in the object

Author(s): Person(s), group(s) or organisation(s) responsible for the intellectual content of the object

Owner(s): Current legal owner(s) of the object

Source(s): The source(s) of the information found in the object

Purpose: Reason why the object was created and intended user communities

Funding & conditions: Source of funding and the conditions set

Subject: Data topics and keywords describing the subject matter of the data

Geographical coverage: The countries, regions, cities etc. covered in the data

Temporal coverage: The date (or date range) covered by the data

Date of collection: The date (or date range) of data collection (may be same as the temporal coverage)

Sample size & description: The number of individuals surveyed and characteristics

Current location: Path/www. Address/physical location where the object can be found

Format: Physical or electronic format of object

Size: Physical or electronic size of object

Restrictions: Access or security restrictions placed object by funder/researcher

Documentation available: e.g. user manuals, code books, references to its location

Retention period: Planned retention period for the data & ideal retention period

Additional notes: