



Citation for published version:

Pink, CJ, Cope, J, Jordan, KM & Jones, K 2013, *Research360: Faculty-Industry Data Requirements Report*. University of Bath, Bath.

Publication date:

2013

Document Version

Early version, also known as pre-print

[Link to publication](#)

Publisher Rights

Unspecified

University of Bath

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Research360

Faculty-Industry Data Requirements Report

University of Bath
July 2013

Compiled by Catherine Pink, Jez Cope, Katy Jordan and Kara Jones
Data gathered by Catherine Pink, Jez Cope, Katy Jordan, Linda Humphreys and Neil Beagrie.

With thanks to those who took the time to complete the online survey, those who generously provided their time and experience for case studies and members of the Digital Curation Centre for help designing the survey.

Contents

1. Executive Summary	3
2. Motivation & Background	5
3. Methodology	6
4. Participation	6
5. Results.....	7
5.1. Nature of research	7
5.2. Funding.....	8
5.3. Data ownership.....	9
5.3.1. Data sources.....	9
5.4. Data management planning	11
5.5. Current data storage – live data	12
5.5.1. Non-digital data.....	12
5.5.2. Data storage	12
5.5.3. Use of BUCS-managed storage space	14
5.5.4. Metadata/description.....	14
5.6. Data loss.....	14
5.6.1. Amount of data lost.....	14
5.6.2. Reasons for loss	15
5.6.3. Loss prevention	16
5.7. Data Sharing.....	17
5.7.1. Amount of data shared.....	17
5.7.2. Restrictions.....	18
5.7.3. Mechanisms of sharing	20
5.7.4. Version control.....	21
5.8. Data publishing and archiving	21
5.8.1. Current projects	21
5.8.2. Completed projects	21
5.8.3. Non-digital data.....	23
5.9. Perceptions.....	24
6. Case Study 1: Pilot Research Project in a Single University Department	27
7. Case Study 2: Cross-Faculty Research Facility.....	29
8. Case Study 3: Sharing Personal Data with UK and International Collaborators	31
9. Management View	34
10. Recommendations	37
11. References.....	39
12. Appendix 1: Survey Questions	40

1. Executive Summary

Research360 is an 18-month Jisc-funded project, running from autumn 2011 to summer 2013 at the University of Bath. The project aims to develop the human and technical infrastructure required to facilitate good management of research data.

This report looks at faculty-industry research data requirements using responses gathered from an online survey, interviews and a synthesis of findings from an independent report on research data management at the University.

We have found that collaborative working, particularly with industry or commercial partners, is common for Bath researchers. Negotiating data ownership with these partners can be a source of frustration, and is likely to impact on what researchers feel they can or should do with the data they are working on throughout the research data lifecycle. The use of research data management plans is not currently widespread, although the early creation of such plans may assist with ownership, storage and archive issues at a later date.

Current research data storage was explored, with sharing of data with industrial collaborators cited as a common problem. This ranged from difficulty using shared drives to slow network connections hindering data exchange. Those with industrial connections were much more likely than average to be sharing larger volumes of data.

There appears to be some uncertainty about sharing current primary research data outside of an ongoing project. Over 70% of respondents either don't want to, aren't allowed to, or don't know whether to make their primary research data available. Free text comments suggest that respondents feel this would be additional workload, would compromise confidentiality or would jeopardise publication.

Respondents who do plan to make their current research data publicly available appear to be approaching this in a variety of ways and again there is a large element of uncertainty over how this will be accomplished. There appears to be a willingness to publish data as supplementary to a scholarly article or in a data journal, as well as in a repository, which is encouraging in terms of preservation and citation capability.

The restrictions that apply to the availability of data are often related to collaborator requirements or the confidential or sensitive nature of the data. The relationship between the researcher and any industrial or commercial collaborators clearly impacts heavily on whether data can be made (or at least perceived to be made) available without restriction.

Survey respondents highlighted the need for clarity relating to data management and ownership, and how industry-imposed restrictions on data use can run counter to the Research Councils' requirements. They look to the University to put measures in place to manage industry's expectations in this regard, perhaps with the use of standardised documents and agreements.

Throughout, the responses demonstrate a need for advocacy and training to raise researchers' levels of awareness and skills in data management. There is a clear expectation by researchers that while data management is primarily their responsibility, the University has a definite role to play here by providing support and training alongside the provision of a data storage and archive system.

Based on the survey results, interviews and an independent report, the following recommendations are suggested:

1. Infrastructure and services should be further developed to support researchers at Bath, particularly with reference to meeting expanding data storage needs, funder requirements for access to data and the requirements of the 'University of Bath Roadmap for EPSRC'. As much of the research undertaken at the University of Bath is collaborative in nature, this infrastructure must include suitable tools to facilitate the sharing of data during the active phase of the project.
2. Training and resources should be enhanced to develop graduates with good data management skills, including an understanding of the issues arising from collaboration and future sharing of data. Good quality training resources are needed both within structured training schedules, and at the point of need. Guidance on current University provision for data management and also on the appropriateness of third-party services must be made readily accessible.
3. Clarification on data ownership issues, particularly when negotiating with commercial partners, should be provided. This will help reduce confusion over what can or can't be done with the data throughout the research lifecycle. This clarification should include both a standard University policy on data management and also signposting for help in case of issues that fall outside of regular policy or procedures, for example when non-disclosure agreements run counter to research council or other funder agreements.
4. Data management plans should be used more widely to focus processes on meeting funding or industry requirements. Templates and supporting guidance should be made readily accessible for use both during the grant application process and by research postgraduate students.
5. An outline of the benefits of good data management, both to the University community and to external partners such as industry and commerce, would provide clarification and support for the activity for all stakeholders.
6. A common thread amongst all themes and responses was the desire for institutional support throughout the research lifecycle. Resource should be invested not only in technical infrastructure but also in professional support. The latter would provide infrastructure development, training researchers in good data management practices, and ensuring compliance with funder and industry requirements.

2. Motivation & Background

Research360 is an 18-month Jisc-funded project, running from autumn 2011 to summer 2013 at the University of Bath. The project aims to develop the human and technical infrastructure required to facilitate management of research data.

The Research360 project arose as a response both to the need for enhanced data management within the University, driven by the Research Data Steering Group, and changing external drivers. Paramount amongst these was the EPSRC Policy Framework on Research Data, published by the EPSRC in early 2011¹. Institutions in receipt of EPSRC funding were required to have a roadmap in place by May 2012, setting out how they plan to put in place the necessary policies, guidance, training and technical infrastructure to ensure compliance with the expectations by May 2015.

In order to develop these new data management policies, guidelines, training and to develop a new technical infrastructure, including an institutional data repository, it is necessary to better understand how research data is currently managed across the institution.

In May 2011, the University of Bath carried out an audit of its data assets using the Jisc funded Data Asset Framework (DAF), which showed that

1. most respondents are satisfied with their data storage, although insufficient storage space is the biggest data management issue. There is an over-reliance on external hard drives for backing-up data. This is a concern for the retrievability and security of data;
2. data management practices are guided by intuition rather than informed by good practice;
3. there is a need for services to share research data during the life of the project with internal and external partners. These should be easy to use, geographically accessible, and employ version control;
4. data reuse and sharing is considered a good idea but very few researchers are currently actively sharing their data, or providing robust metadata for archival storage;
5. guidance on, and storage for the preservation of data would be appreciated².

This report builds on that survey, with particular reference to specific data management needs and the requirements of research conducted in collaboration with industry and commerce.

¹ EPSRC Policy Framework on Research Data. Available: <http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx>. Accessed 6 May 2013.

² Jones, K., 2011. *Assessing institutional data storage and management using the Data Asset Framework (DAF) methodology at the University of Bath*. Bath: University of Bath. Available: <http://opus.bath.ac.uk/24960/>. Accessed: 20 November 2012.

Current research data management infrastructure

The current research support infrastructure at the University of Bath is distributed between the Library, Bath University Computing Services (BUCS), the Research Development and Support Office (RDSO), and individual research groups and facilities.

BUCS run the majority of central IT services. The University's institutional repository, Opus, is a library EPrints system for published research outputs. It contains a mix of 22,000 metadata-only and full text records. Opus already hosts some research data, for architecture research projects funded by AHRC, taken on when AHDS closed.

The University uses Pure as its Current Research Information System (CRIS). Pure is integrated with the HR and finance systems, the student system and the institutional publications repository (Opus). Pure enables research grants to be linked to publications in order to report on outputs to the Research Councils, for REF, to Heads of Departments and others.

3. Methodology

The structure and content of the Research Data Survey drew on both the Digital Curation Centre's CARDIO tool and the Digital Asset Framework report carried out in May 2011 at the University of Bath. The Research 360 project team scrutinised the CARDIO tool and extracted what it judged to be the key elements, structuring a shorter and simpler survey which would be as quick and easy as possible for participants to complete. The DAF report was consulted to ensure that the survey focused on areas previously identified as problematic for the University of Bath.

The survey was constructed using the Bristol Online Surveys resource. The first iteration was reviewed by the Research360 project team and further refined. The Digital Curation Centre (DCC) Institutional Support Officers for the University of Bath were also consulted and suggested further refinements.

Results were analysed according to whether research was being undertaken in association with industry. Using the filter option in Bristol Online Surveys, responses were filtered based on whether the answer to question 4 'Who funds your chosen research project?' was 'industrial/commercial sponsor' or whether the response to question 5 'who are you working with on your chosen research project?' was 'collaborators in commercial industry'.

4. Participation

The survey was launched on June 20th 2012 and ran for 7 weeks until August 7th 2012. The University was running a number of other surveys at this time, and its policy on promotion of surveys requires that it be low-key and non-aggressive. The Research360 project team therefore judged that a relatively lengthy time period would allow participants adequate time to discover and complete the survey. Promotion of the survey included an email to all researchers followed up by a reminder email cascaded via Associate Deans for Research. An internal news report was published when the survey was launched and an item included in the weekly staff newsletter. In addition, a banner was added to the BUCS homepage for the duration of the survey. An added inducement to participation was the opportunity for all

those who completed the survey to opt into a draw to win a Kindle 3G, eventually awarded via a random draw to a postgraduate research student from the School of Management.

In the event, the survey attracted 210 respondents out of the target demographic of 6041. This constitutes a response rate of 3.5%, which was low and the results should be considered in that context. However, respondents were drawn from all academic departments and faculties, and from all relevant researcher roles (Figure 1). Nearly 27% of respondents were Principal Investigators, a further 25% were research postgraduates, 20% were taught postgraduates while research officers constituted nearly 14% of the total. Their responses proved to be wide-ranging and detailed, yielding a broad picture of all aspects of research data management and storage at the University of Bath.

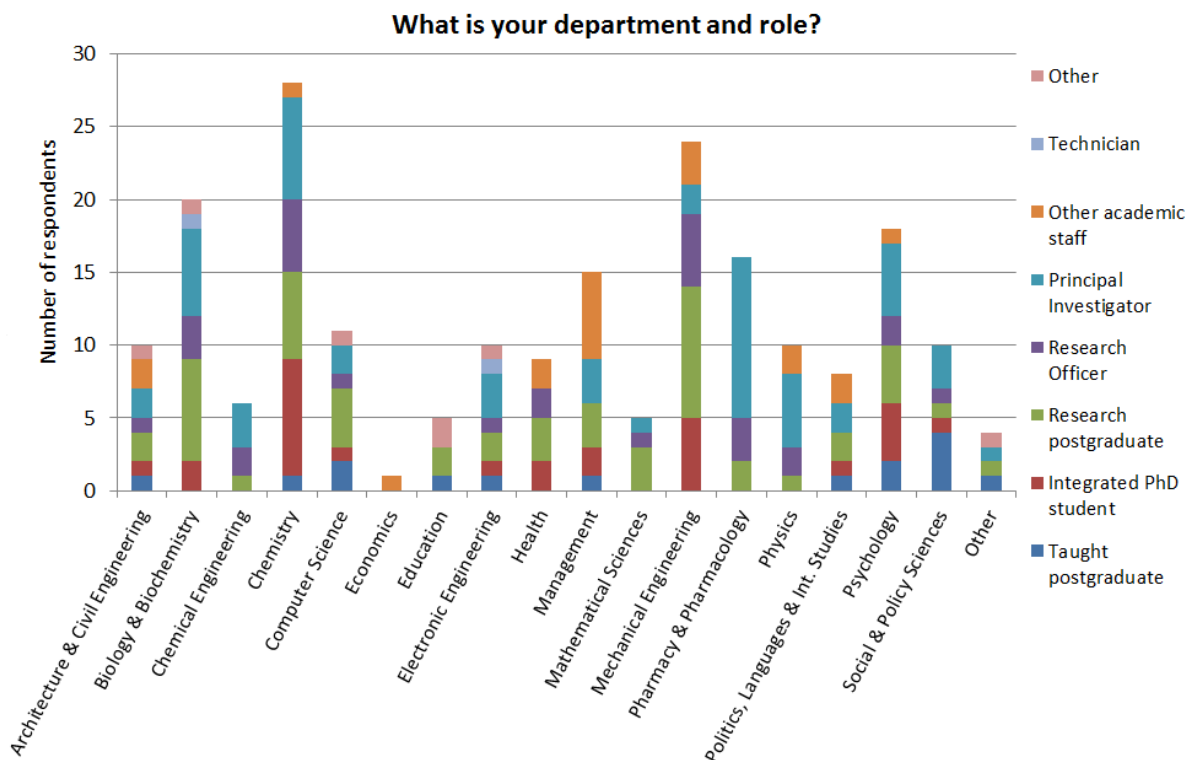


Figure 1: Demographic of respondents showing that all departments and researcher roles were represented.

5. Results

For a number of questions, researchers were asked to answer with regard to their most recent or main current research project. This is referred to throughout as their “chosen research project”.

5.1. Nature of research

The majority of postgraduate students are currently working on only one research project. However, research officers, PIs and other academic staff tend to work on multiple projects at any given time (Figure 2).



Figure 2: The number of projects a researcher is currently working on split by the type of researcher.

5.2. Funding

A large number of respondents indicated that their research was funded either by the University or by a single UK funding body such as a Research Council (Figure 3).

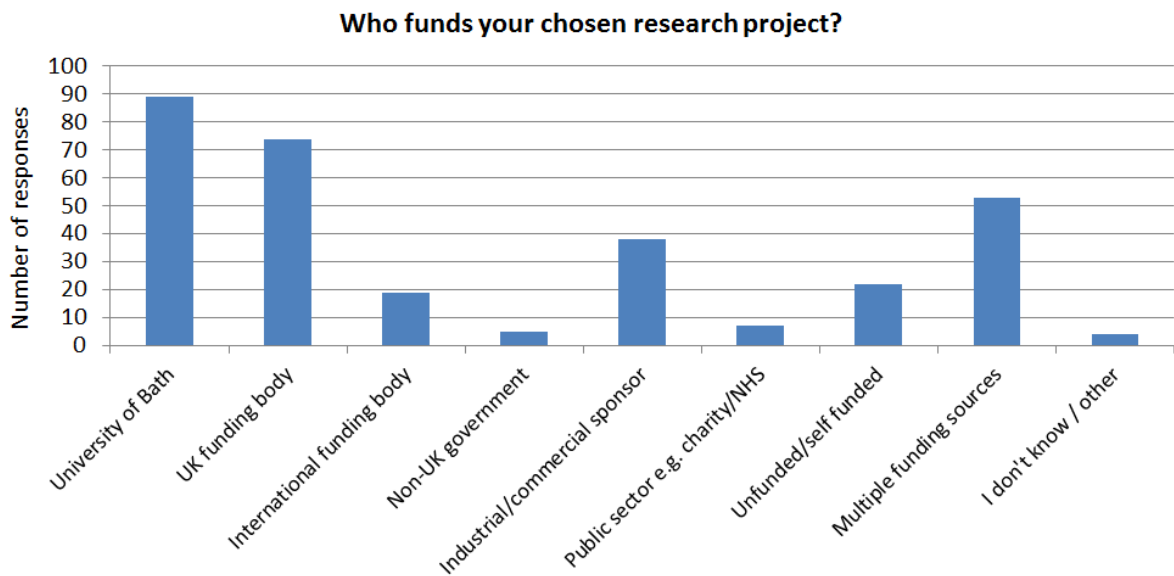


Figure 3: Researchers were asked to select all applicable sources of research funding for their chosen research project.

The relatively high number of unfunded or self-funded students possibly reflects the number of masters students who participated in the survey, although some pilot studies undertaken to support grant applications also fell into this category. University of Bath funding was selected predominantly by postgraduate researchers on University-funded studentships, and Principle Investigators employed by the University but also in receipt of external funding.

Non-UK government funding reflected international students either sponsored by home governments or visiting from international universities.

There is a clear need for University policy to guide the data management activities of these researchers, where no external policy would apply. Where University funding coincides with sponsorship from UK funding bodies such as the Research Councils or Wellcome Trust, University policies and guidance should align with external requirements.

Collaboration

The survey asked researchers who they were working with on their research projects. Some research is carried out independently or under the direction of a single supervisor only. However, the majority of research undertaken at the University of Bath is collaborative in nature, with many research partners based outside the University at other universities; in industry or the public sector; both nationally and outside the UK (Figure 4).

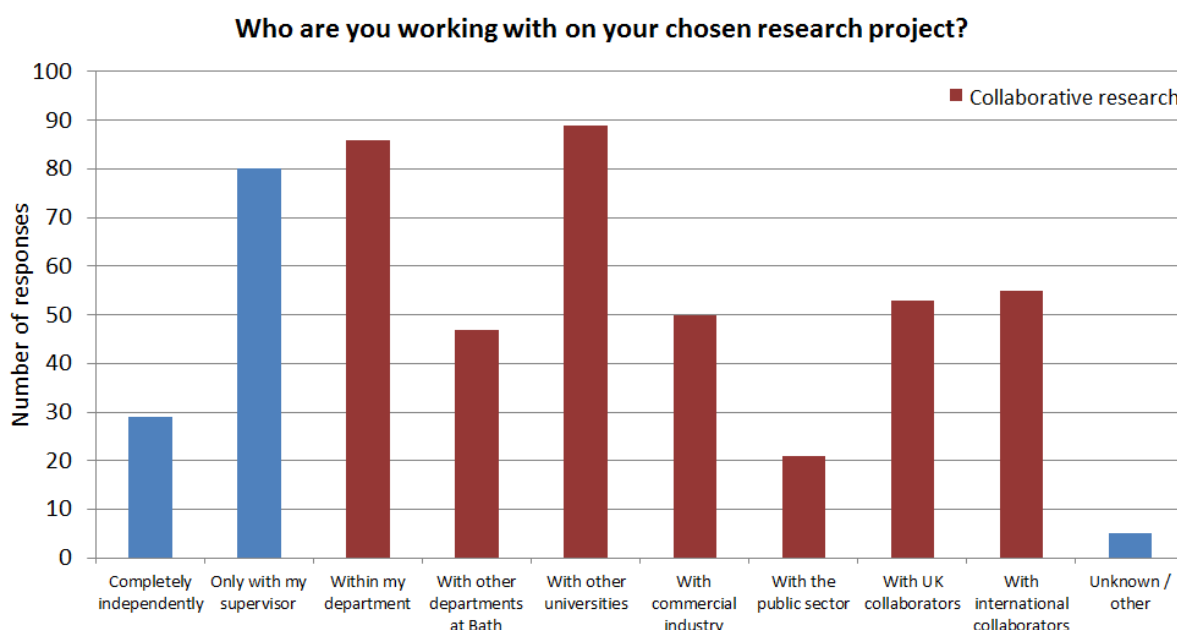


Figure 4: Researchers were asked to select all applicable partners for their chosen research project.

The collaborative nature of research demonstrates the need for suitable tools to facilitate sharing of data within dispersed research groups during the active phase of a research project. i.e. prior to publication and data archive.

5.3. Data ownership

5.3.1. Data sources

Most researchers generate or collect their own research data. However, many also indicated that they used data from other sources including colleagues, data repositories, commercial or other data providers.

Within this context, it is important to consider the range of formats of data. For example, researchers who build on theorems as their primary research material or who develop theoretical/mathematical models: whilst they do not consider their primary research material

to be data, their research outputs could be managed as data – stored securely, shared and preserved for the future.

Data Management

The management of data was described as ‘ensuring its quality and security of storage, describing, archiving and publishing it’. Whilst the majority of researchers felt that they were responsible for management of their research data, some still felt that the onus of responsibility fell on supervisors, designated data stewards or the University (Figure 5). A few general comments in the free text box suggest that some researchers do not know what arrangements exist, for example:

“Currently the University offers no help on this front, no guidance, no help”

This suggests a need for clarification and communication on data management, particularly in terms of University offerings.

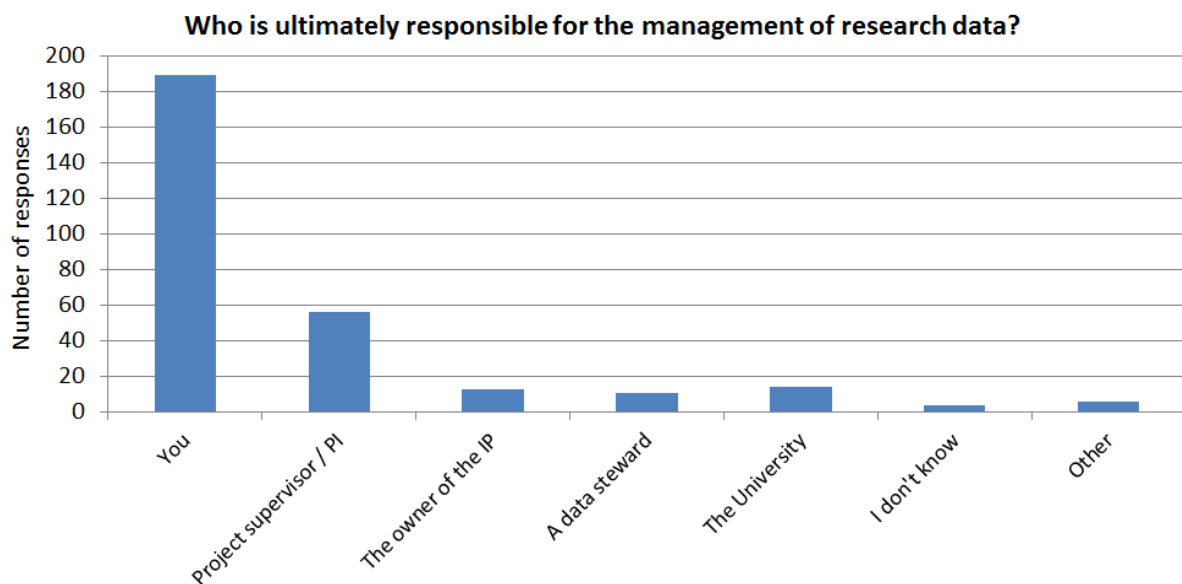


Figure 5: Researchers were asked who they thought was responsible for management of researcher data. Respondents were allowed to select more than one option.

Researchers expressed frustrations in negotiating and managing data ownership with commercial partners, suggesting that the extent to which a commercial partner is a collaborator in the research process can be challenging.

General comments in the free text box for this section suggest that there is some confusion over data ownership. This is important to know as data ownership is likely to impact what researchers think they can or should do with the data they are working on throughout the research lifecycle.

“My research is owned by me for the duration of my PhD”

“Ownership could depend on the type of product”

“I’m not sure exactly what is meant by ‘ownership’”

“Coding is owned by me/the research team”

5.4. Data management planning

Data Management Plans

Of the survey respondents, >81% had not produced a data management plan. Only 8.3% of respondents had voluntarily produced a data management plan. The remaining 10.3% had produced a plan because they were required to do so, usually by funders or as part of an ethics approval process to demonstrate confidentiality and for assurance that sensitive data will be managed appropriately (Figure 6).

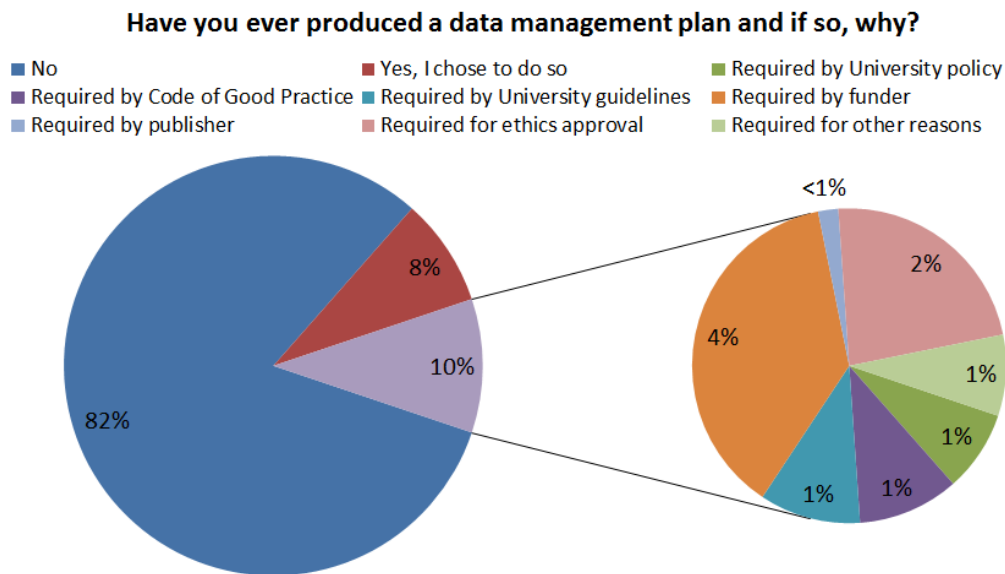


Figure 6: Researchers were asked if they had ever produced a data management plan. Those who had been required to do so were asked where the requirement had come from.

Those respondents who had produced a data management plan often had not used a template (57.1%). Others based their work on guidance provided by the funding body as part of the application form, or based on templates already existing within their research group. There has been very little use to date of any DMPonline tools for creating data management plans, and it would perhaps be useful to know if those respondents who did not use a template managed to satisfy any data management requirements expected of them.

Use of a Data Management Plan

Half of the respondents who had developed a data management plan did in fact refer to it after producing it, finding it useful in managing their data. However a substantial proportion (45%) did not refer to it again, or did not find it useful. This may suggest that the plan was not designed effectively, or that the respondents did not see or understand the relationship between the plan and implications for actual data activities during/following their research.

5.5. Current data storage – live data

5.5.1. Non-digital data

Nearly two-thirds of the university’s researchers store some of their research data in “non-digital” forms, with only 35.2% of respondents storing all of their data in digital form (Figure 7). Of those with non-digital data, the vast majority is stored in paper forms, such as lab notebooks or interview notes. That so much data is non-digital has long term implications for archive and preservation.

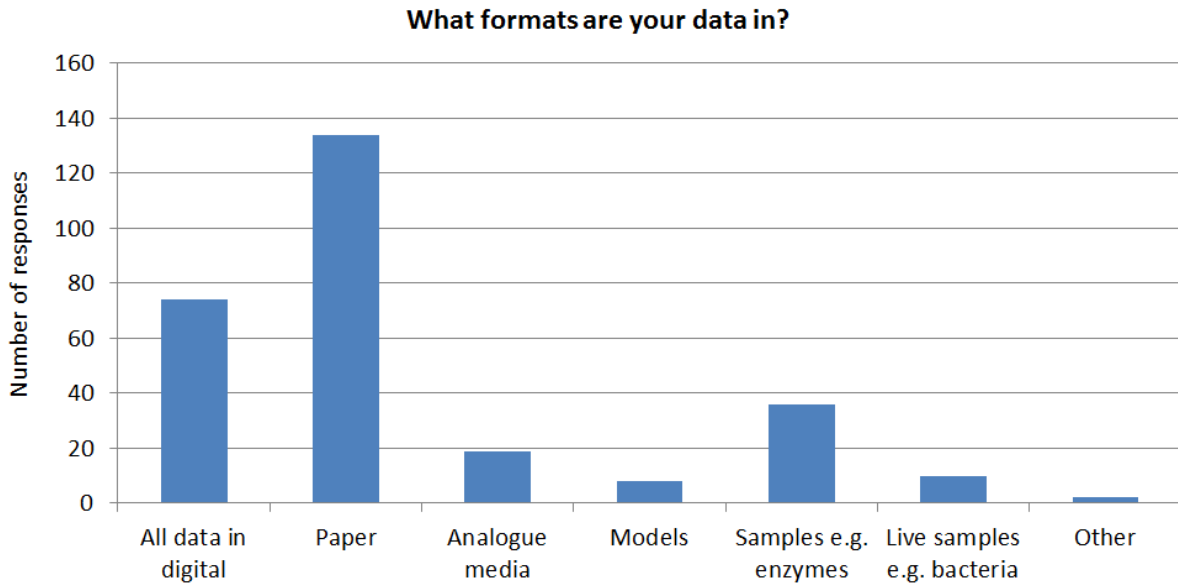


Figure 7: Researchers were asked to select all formats that their data was in for their chosen research project.

Analogue media (such as tapes), physical models and samples were also regarded as data by some respondents. One respondent specifically stated a need to retain physical gels to check against digitised copies due to potential faults in the scanning process, although it is not known how long this storage was typically required for.

5.5.2. Data storage

We can regard storage of data on the X: (research) or H: (personal) drives, or other university managed storage, as a “gold standard” for resilience. By this measure, 68.1% (143 respondents; Figure 8) store at least some of their data in a safe, resilient form.

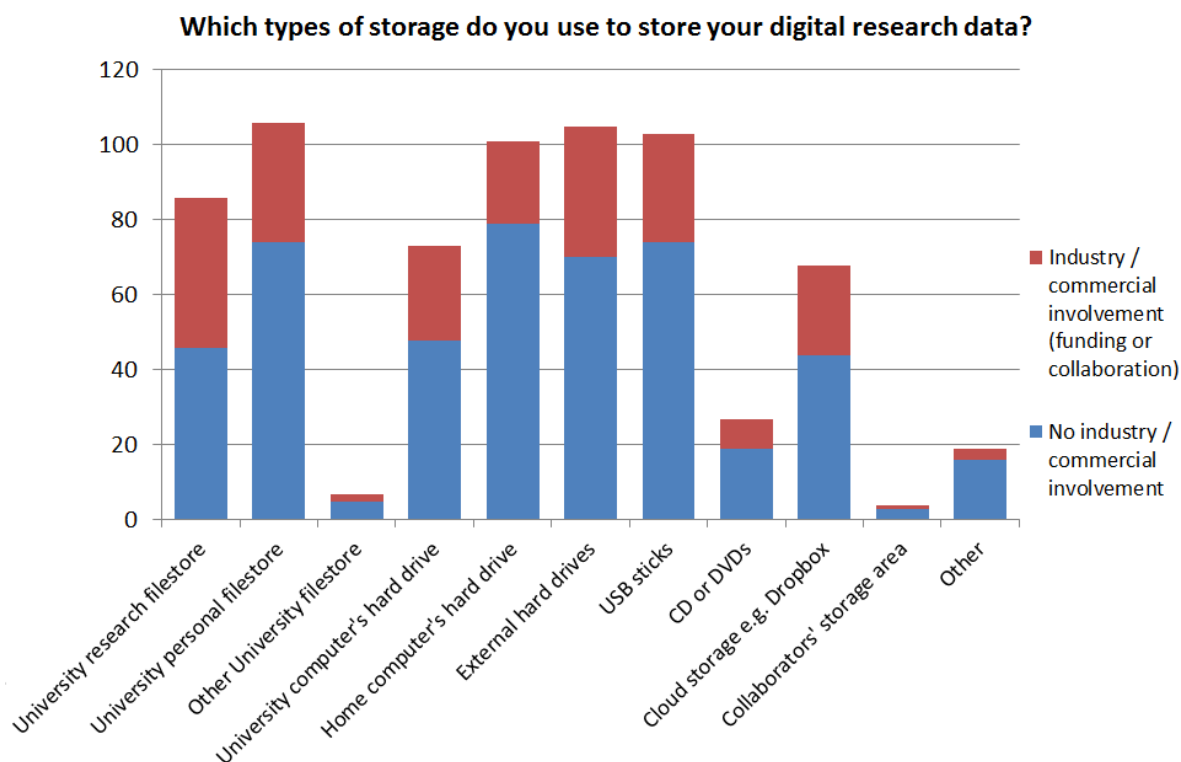


Figure 8: Researchers were asked to select all locations where they store digital data for their chosen research project.

Some departments are maintaining their own storage systems: in particular, several respondents from Computer Science reported storing their data on departmental servers and databases. One respondent reported using their own separate storage and backup system (in addition to central storage) in order to overcome concerns over performance of the networked storage. Another respondent stores data on a hospital server, which may be a data protection requirement.

In the free text comments, it emerged that a common pattern was to keep a working copy on a local desktop or laptop computer, and use networked storage as “backup”. This is fine as long as the “backup” takes place regularly, but it would be preferable to work from the network copy whenever possible. For this to be commonly accepted, it needs to be easy (and ideally automatic) to synchronise working files between the network and local drive when a user wishes to work offline.

Nearly a third of respondents (68 of 210) reported using Dropbox or another online service to store their research data. This indicates a clear need for policy or guidance on when this service is and is not appropriate for research data storage and sharing.

Other forms of data storage mentioned include:

- SourceForge (a public site which hosts open-source software) — probably fine as long as this isn't the only place where current software code is stored;
- Learning Materials Filestore (developed internally by Web Services for sharing large files) — only intended for sharing files with other Bath users, and not suitable for long-term storage.

5.5.3. Use of BUCS-managed storage space

The question “Are there any reasons why you would not use storage space managed by the University of Bath?” was phrased to solicit barriers and as such, responses may have drawn out particularly contentious issues.

A number of genuine challenges were posed that may necessitate a change either to the services offered or to the way in which existing services are offered. By far the most common responses referred to:

- Routine off-campus working — access to shared drives via VPN over the internet is slow and unreliable;
- Routine mobile working — shared drives cannot be accessed at all without an internet connection;
- A need to share data with external collaborators, who are unable to get access to shared drives.

In addition, many of the free-text answers highlighted potential misconceptions and unrealistic expectations of the service. These may have technological solutions in some cases, but could also be dealt with through appropriate communication with researchers.

5.5.4. Metadata/description

The majority of respondents (76.4%) describe their data using strategies that often prove problematic in the long term (filenames: 60.0%; column names: 16.4%). Only 17% use the robust strategies of directly attaching metadata (6.7%) or storing metadata/descriptions in a separate file alongside the data (10.3%). Further, 6.7% of respondents do not keep records of what their data is, which would present a considerable barrier to re-use, both by themselves and others.

Overall, respondents had a high degree of confidence in their description of their data, with 88.1% feeling (Agree/Strongly agree) that they would be able to interpret their data in 3 years' time (with only 2.5% disagreeing and none strongly disagreeing), and 60.2% feeling that others could interpret their data well enough to reuse it.

An important minority of respondents (16.4% Disagree/Strongly disagree) felt that other people would be unable to interpret their research data appropriately, with a further 23.4% neither agreeing nor disagreeing. Again, this would have longer-term implications for data archive and re-use.

5.6. Data loss

5.6.1. Amount of data lost

Overall, 76.1% of respondents reported never having lost any research data (Figure 9). This subsection and the next (“Reasons for loss”) refer only to those 23.9% who have lost data, and percentages should be interpreted in that context. It is also important to note that these results are unlikely to be representative of the University as a whole. Respondents may have

been a self-selecting group, with a slight bias towards researchers who have previously lost data and subsequently have a greater interest in data management issues.

Of the 47 researchers who had lost research data, just over half (55.9%) had only done so once, with a further 30.5% having lost data up to 3 times (Figure 9). Two respondents have lost data frequently, with one losing data around once a month and one losing data even more frequently than that, though these two respondents have not lost more than a few days-worth of work in total.

The majority of respondents who had reported losing data have lost only a few days' work in total (71.9%). However, 12.3% have lost around a month's work, and 10.6% have lost considerably more than this (Figure 9). These figures suggest that training to improve data storage and organisation would be worthwhile, relative to the potential costs associated with reproducing lost data.

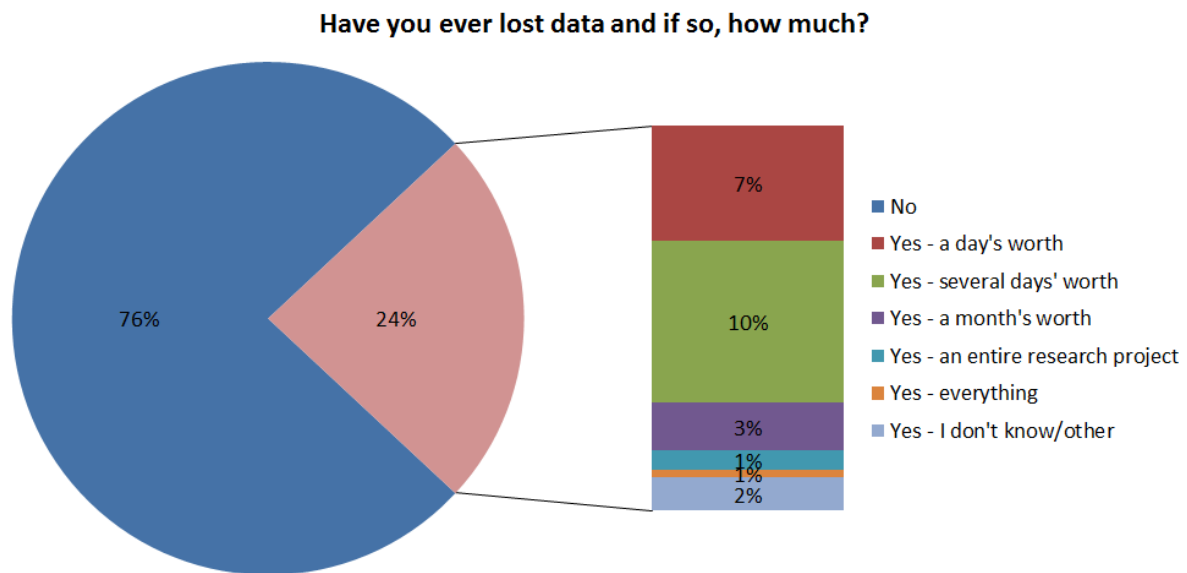


Figure 9: Researchers who had lost data were asked how much they typically lost. Note that percentages shown on the plot relate to the total responses whereas those in the text relate only to respondents who have lost data.

5.6.2. Reasons for loss

The survey then explored the reasons why data had been lost. All of the possible multiple choice options here were covered at least once, with the top 3 being: hardware failure (48.9%); accidental deletion (31.9%); and file corruption (23.4%).

A number of interesting cases were also mentioned in the free-text comments, including:

- Project or PhD students leaving and taking data with them;
- Loss of power supply to recording instrument;
- Fire damage to samples and paper-format data;
- Spreading data across several storage locations, leading to confusion as to what was stored where.

5.6.3. Loss prevention

Although 68.1% of respondents reported using University-managed network storage (see section 5.5.2 on Data Storage), only 39.0% report this as a means to avoid data loss, which may indicate some lack of understanding of the managed storage.

The survey also highlighted a number of respondents relying on poor forms of backup, such as an external hard drive (58.6%), a second computer (28.1%), or even a second copy on the same computer (42.4%). This last option is no safer than a single copy, since both copies are at equal risk.

However, these individual figures do not tell the whole story: A researcher who uses several of these methods can be regarded as having a resilient strategy, even though individually the methods are not sufficiently resilient. A more detailed analysis was therefore performed to assess the number of respondents using a sufficiently resilient strategy. Possible answers to Question 26 of the survey (Appendix 1) were scored on a scale from 0 to 3, with higher scores representing more resilient options (Table 1). Respondents were assigned a total score based on their responses, with 3 considered the threshold for “safe” backup strategy. By this measure, 52.9% of respondents are using a sufficiently resilient backup strategy, leaving 47.1% still in a precarious situation.

Option	Score
None -- my data could be easily replaced if it were lost:	3
Utilisation of University of Bath managed storage e.g. X:drive:	3
Backed up on an external hard drive:	1
A second copy on my own computer:	0
Additional copies on a second computer:	1
An additional version on an alternative media type e.g. USB stick, CDs:	1
Use of cloud storage e.g. Dropbox, Amazon S3:	0
Scanned copies of non-digital paper media:	1
None of the above:	0
Threshold for “safe” backup strategy	3

Table 1: Scoring system used to measure the resilience of respondents’ backup strategies. Higher scores represent more resilient options, with total scores of 3 or more considered the threshold for a ‘safe’ backup strategy

26.2% of respondents reported using cloud storage, such as Dropbox or Amazon S3. Although this is potentially highly resilient, it places responsibility for that resilience on a third party with whom the University has no formal Service Level Agreement. In many cases, it

also places data in legal jurisdictions such as the US which may, for example, be incompatible with EU data protection law.

A small number of respondents (3.3%) reported that their data could easily be replaced if lost and that they therefore have no precautions to avoid data loss.

Some good and bad practices were highlighted in the free-text comments. The good include:

“All lab notebooks ... completed in duplicate, using carbon paper ... copies are stored in different locations”;

“Weekly disk mirror and occasional LTO tape archives ... both on and off site”;

“Utilisation of University of Bristol managed storage”.

The bad include:

“Email docs to myself frequently”.

One respondent (to Question 31 on version tracking, rather than this question specifically) reports relying “on Time Machine with a dual drive external enclosure in RAID0 mode”. This is worrying as RAID0 provides excellent access speed but *increases* risk of data loss from drive failure, which is the opposite of what is required of a backup drive.

5.7. Data Sharing

5.7.1. Amount of data shared

There was a broad spread of volume of data shared with others, with 22.8% of respondents sharing less than 10MB, 22.2% sharing 10-100MB and 20.0% sharing 100MB-1GB of data. There is a need to support sharing of large amounts of data, with 15.6% of respondents sharing 1GB-1TB, and 2.8% sharing more than 1TB of data.

Those with industrial connections were much more likely to be sharing larger volumes of data, with 26.8% sharing 100MB-1GB, 25.0% 1GB-1TB and 3.6% more than 1TB (against 16.9%, 11.3% and 2.4% respectively for those without industrial connections; Figure 10).

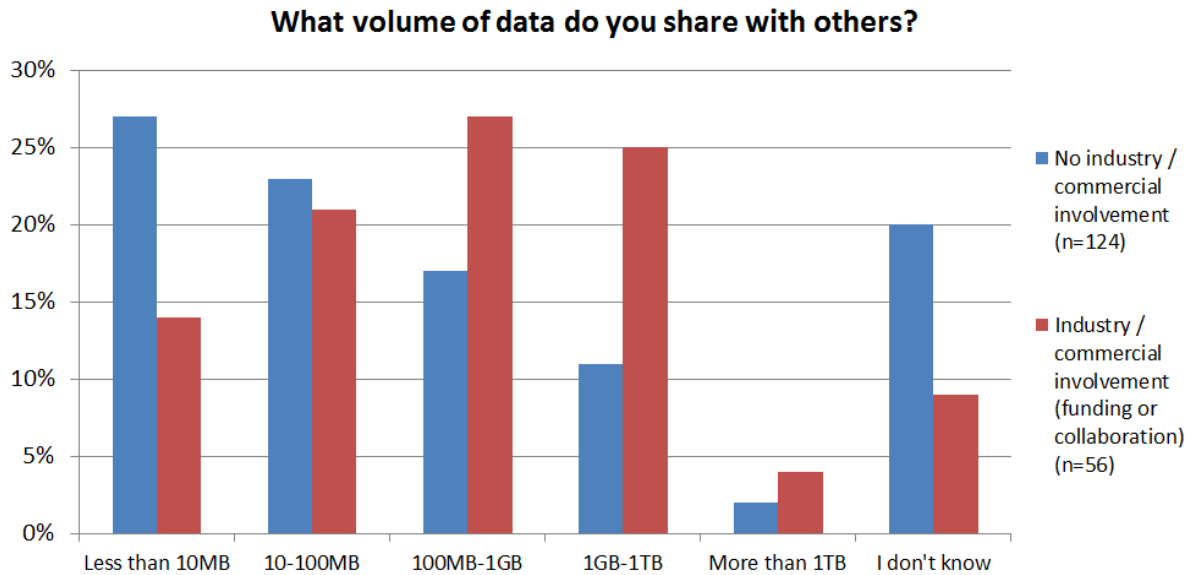


Figure 10: Researchers were asked, for their chosen research project, what volume of data they were sharing with others.

Interestingly, 16.7% of all respondents answered “I don’t know” to this question. This may be because they don’t know how to find out, or just don’t think about file sizes when sharing data.

5.7.2. Restrictions

The majority of respondents (53.7%) thought that all of their data could be shared with other people directly involved in their research project. A further 36.0% were able to share some but not all of their data (specific elements withheld: 15.3%; only what is required by collaborator: 20.7%).

Fewer of those with industrial connections felt that they could not share any of their research data with their collaborators (3.3%, vs. 8.5%; Figure 11) and fewer (49.2% vs. 55.6%; Figure 11) felt able to share all data. More respondents with industrial connections felt able to share a restricted set of data (specific elements withheld and only what is required by collaborator: 44.3% vs. 32.4%; Figure 11). This may reflect the fact that researchers with industrial connections are more likely to have to share some of their data with their partners as part of collaboration agreements.

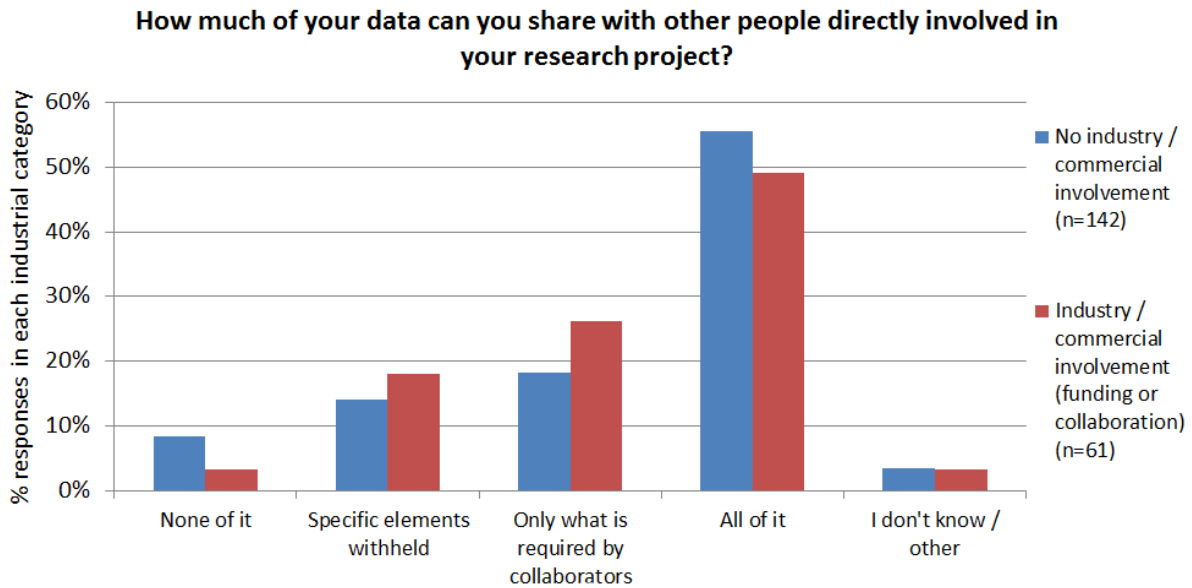


Figure 11: Researchers were asked, for their chosen research project, who they could share their data with. Researchers with industrial connections were more likely to share a restricted set of data.

When asked about restrictions on sharing of data, three main themes emerged. Firstly, many respondents referred to requirements for anonymisation of sensitive data, including pseudonymising participants, stripping personal details and changing dates to “days into the study from time zero”. One researcher referred to a requirement to “get ethical clearance from SREAP [the ethics committee for health research at Bath] to allow [their] supervisors to look at [their] data.” Another made explicit reference to a “code-breaker”, kept securely, which allows reversing of pseudonyms on the rare occasions this is necessary.

A second common theme was non-disclosure agreements (NDAs), which generally restrict access to named individuals within the University, e.g. “Access to folders is limited according to non-disclosure agreement.” One respondent specifically stated that although “due to my NDA, I would not be able to give data from my industrial partner to all my collaborators, however this *does not restrict me from sharing any of my own acquired data thus*” [our emphasis].

The third major theme was the difficulty of giving external collaborators access to the University’s shared research storage area (commonly known as the X:drive), both to pass data to and receive data from those collaborators: “*Need to find a way to share outside university, and also import temporarily a large amount of data.*” One example cited their commercial partner’s IT policy preventing the use of VPN to access the University’s infrastructure.

Other relevant comments included:

“Might have restrictions in the future”;

“Amount of data is going up hugely”;

“Only under terms of license as directed by head of group”;

“The data which I share with others are just articles, chapters of books and academic published data.”

5.7.3. Mechanisms of sharing

The majority of respondents (61.4%) reported using email as a method for sharing data. This is a concern as email is inherently insecure, passing through many computers on the way to its destination. Third-party services such as Dropbox (27.1% of respondents) narrowly edge out University shared storage (26.7% of respondents) as a mechanism by which data is shared, giving further cause for concern. A small number of all respondents reported making use of a University web service such as their personal web page (7.6%) or a personal (non-university) website (1.4%).

Use of physical digital media (40.0% of all respondents) and paper copies (20.0% of all respondents) are more secure ways of sharing data, as long as the physical security of the medium is guaranteed (i.e. it is not lost or stolen). Several respondents (5.2% of all respondents) reported using a service provided by a collaborator.

Where researchers had industrial connections (64 of 210 respondents), they were more likely to make use of physical digital media (50.0% vs. 35.6%; Figure 12), Dropbox or similar cloud tools (37.5% vs. 22.6%; Figure 12) and tools provided by a collaborator (10.9% vs. 2.7%; Figure 12) than their colleagues without industrial connections.

Comments in the free text box referenced other methods of data sharing, including Sourceforge and a personal server maintained outside the University. There was also a request for SFTP, something that has repeatedly been raised by researchers as part of support provided via research-data@bath.ac.uk:

“I would like an easy to use FTP (or SFTP) site.”

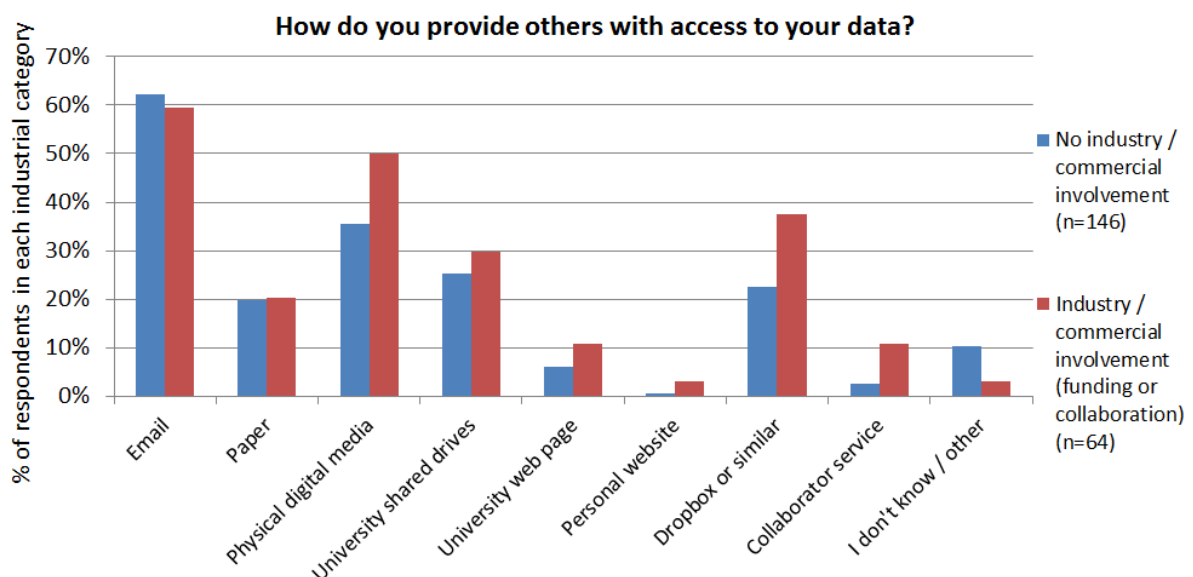


Figure 12: Researchers were asked to select all methods by which they were sharing data for their chosen research project. Researchers with industrial connections were more likely to make use of physical digital media, cloud or collaborator services.

5.7.4. Version control

Two thirds (66.6%) of all respondents reported using file- or folder-naming conventions to track different versions of files – a simple yet effective strategy if used sensibly. 21.4% reported not saving different versions of files at all, instead overwriting them as they are updated.

A small number (8.1% of all respondents) reported using automated backup software, such as Time Machine, which saves old versions of files, and 6.2% use dedicated version control software. No respondents reported using a document management system such as SharePoint, which is unsurprising as these systems are often difficult to set up and maintain, and no such system is currently provided by the university. Two free-text answers mentioned *“use of wiki software to produce documents, which provides lots of version control.”*

5.8. Data publishing and archiving

5.8.1. Current projects

There appeared to be some uncertainty towards current primary research data sharing. Over 70% of respondents either don't want to, aren't allowed to or don't know whether to make their primary research data available. Free text comments suggested that respondents feel this to be additional workload or would compromise confidentiality or jeopardise publication.

“No, because that's unnecessary work. If anyone asks for it I will happily give them whatever they want, subject to concerns about commercial exploitability, but I'm not going to spend time sticking it up somewhere so nobody can ever look at it!!”

“I would like to as I think it creates transparency in reporting. However, I don't have the ethical clearance to do so and it is rarely done in research with psychological variables. I'm not sure how it would work”

Respondents who did plan to make their research data publicly available appeared to be approaching this in a variety of ways and again there was a large element of uncertainty over how this would be accomplished. There appeared to be a willingness to publish data as supplementary material accompanying a scholarly article or in a data journal, as well as in a repository, which is encouraging in terms of preservation and citation capability.

The restrictions that researchers felt applied to the availability of data were often related to collaborator requirements or the confidential or sensitive nature of the data. Anonymity of names, locations and other identifiable information was a significant motivation to restrict data availability. The future commercialisation and patent potential of data was also a disincentive to sharing. The relationship between the researcher and any industrial or commercial collaborators clearly impacted heavily on whether data can be made (or at least perceived to be made) available without restriction.

5.8.2. Completed projects

A large proportion (89.3%) of all respondents had kept research data from completed projects. This data tended to be a reasonable size, although most could be kept on portable

storage devices or hard drives. A connection with industry tended to increase the amount of data retained (Figure 13).

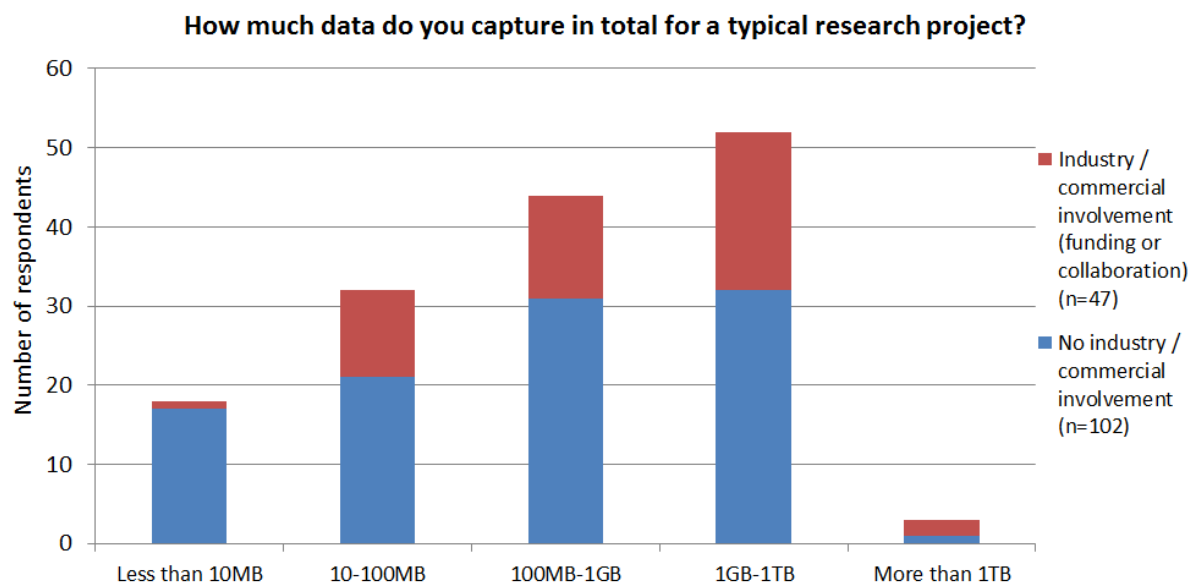


Figure 13: Researchers who had completed a project were asked how much total data was captured for a typical project.

Making research data from completed projects publicly available was not an activity undertaken by most respondents. Of those who had done so, a substantial number had made data available as supplementary material for scholarly articles.

“Published data in science journals available but not the primary unanalyzed data”

“Data published in papers”

If data had been made publicly available and had been re-used, this tended to be by the researcher for different projects, or by different research groups, and the creator was generally given credit for the work. The majority of researchers had not been asked for their unpublished research data, but those who had usually received the request via informal channels. Responses indicated there was uncertainty over whether work had actually been re-used.

Responses in the free-text box for this section indicated a higher degree of understanding for retaining and managing consent forms about collected data. Adhering to guidelines and requirements for ethics approvals seemed to be well embedded in the actions of researchers. This may be an area of understanding that can be transferred to research data management, in terms of thinking about processes and procedures before starting research.

“We have to keep human participant consent forms for 5 years. These are usually paper, for signatures.”

“Not by law, but by international guidelines - 5 years from study completion”

“Yes for a period of two-three years (depending on the project) upon which they can be destroyed.”

“Yes, consent forms have to be kept (separately from collated, anonymised data). All data needs to be destroyed ten years after end of project.”

“Yes, under some format (unspecified), generally in the field for 3-5 years.”

“I am required to keep records for five years after the end of the study as laid out in the ethics application.”

“Yes, under BPS guidelines I should keep all consent forms for at least 5 years (or until data has been disposed of)”

5.8.3. Non-digital data

Most respondents had an idea of the length of time for which they should keep their non-digital research data. 14.1% (of all respondents) replied they did not need to keep non-digital data at all, whilst 19.2% did not know (Figure 14). This may indicate that training on research data management should extend to non-digital data as there is a knowledge gap in this area.

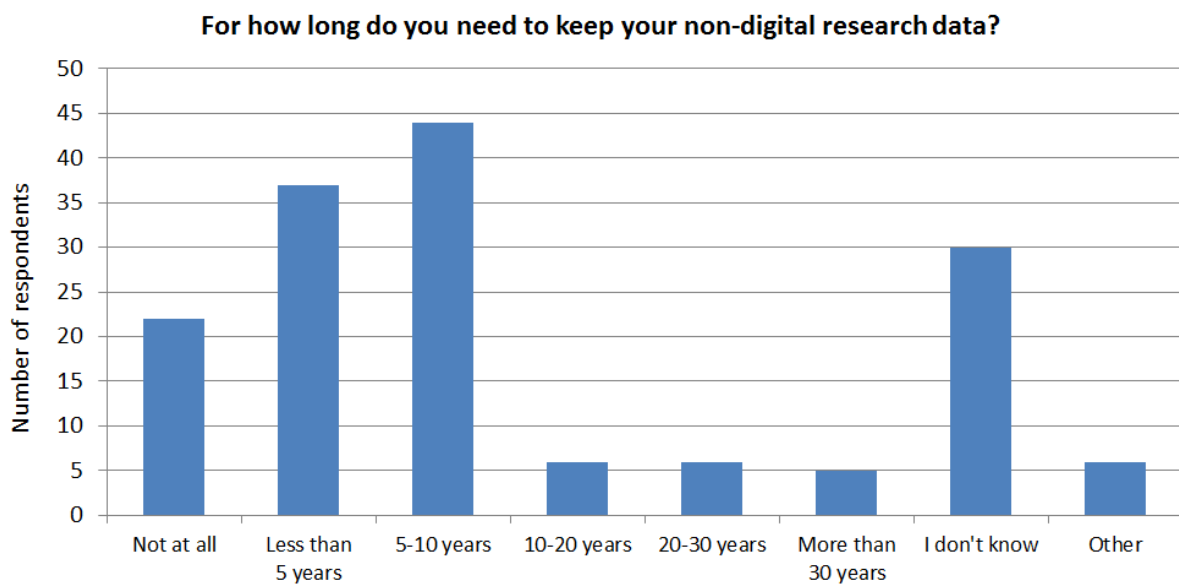


Figure 14: Researchers were asked for how long they needed to keep their non-digital data.

Most of the non-digital data referred to was compact in size, being less than a filing cabinet, or more likely a medium-sized box (Figure 15). However, it is worth noting that based on the survey respondents alone, this equates to over 70 boxes of data, 36 filing cabinets and over 30 cupboards. If improved storage is to be provided, then it is worth considering both the number of researchers with at least a box of non-digital data and the modal length of time that this data should be retained.

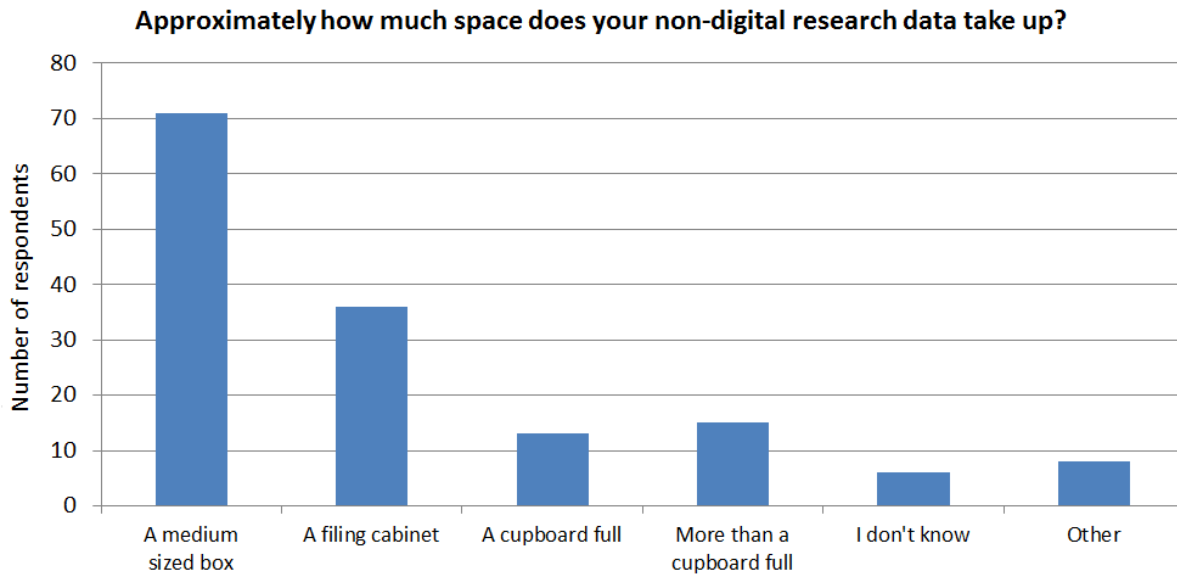


Figure 15: Researchers were asked approximately how much space their non-digital research data takes up.

5.9. Perceptions

The final section of the survey sought to gain an understanding of researchers' opinions about various fundamental issues relating to research data management and accessibility. Typical comments from respondents illustrate the statistical responses in this section (Figure 16).

“If the university doesn't have a policy regarding research data storage and management, it should have – and it needs to be strictly enforced. It would reflect poorly on the university if data is lost.”

There was overwhelming agreement (nearly 95% of respondents) that research data management is primarily the responsibility of the data owner. In addition there was a strong level of recognition (over 47% of respondents) that the University also has a responsibility for ensuring good research data management. However, a substantial minority (nearly 20%) of respondents did not agree that the University had any responsibility in this regard.

“The responsibility should be shared... the University should offer a user-friendly and non-bureaucratic system for researchers.”

Over 55% of respondents felt that publicly-funded research data should be accessible to the public, while 29% were ambivalent on this.

“Publicly funded research data should be available to other academics – not so sure about the public and not commercial organisations.”

Much stronger preferences were displayed regarding intellectual property rights in research data. Over 65% of respondents wanted recognition for sharing their research data; while over 88% felt they should receive recognition when other people re-used their research data.

Respondents were asked to comment on their perceptions of working with industry and other external organisations. One respondent described the difficulties of obtaining data from industry as:

“...often depending on the development of trust relationships over extended periods of time. Anything that might undermine that trust (such as data sharing requirements) would make acquiring data quite impossible.”

Another respondent highlighted the current lack of clarity relating to data management and ownership, and how industry-imposed restrictions on data use can run counter to the research councils’ requirements. They looked to the University to put measures in place to manage industry’s expectations in this regard:

“It may be helpful to ensure that there is a standard legal document that the university uses, or specific clauses to add to company NDAs to include references to research data management. Applicable points here would be who is in charge of managing the data, who owns the data, and the boundaries imposed upon the researcher by the company which inhibit them fulfilling their obligations to the research council.”

Moreover it is not just industry that complicates data management within the University. Sometimes the requirements of data repositories funded by research councils themselves can set up barriers to effective use of research data:

“I have issues with requests from ESDS³ to delete data at the end of the research project when we had not finished writing articles.”

The remainder of this section dealt with perceptions of the importance of research data management and the amount of organisational support needed by researchers. Over 88% of respondents agreed or strongly agreed that research data management is important. Over 51% of respondents wanted to find out more about research data management, and nearly 40% wanted additional support to help them manage their data.

“I have been on a data management training course, but would, at certain stages of the project, appreciate being able to talk with a named person (face-to-face not simply by e-mail!) to ask for individual advice and training specific to my project.”

Over 45% would use the University’s data repository when it is created; however a considerable proportion of respondents (over 36%) were uncertain about this, and a minority (over 17%) stated that they would not use a University data repository. This last group appears to correspond to the nearly 20% of respondents who disagreed that the University has any responsibility for managing research data. It might also demonstrate a lack of understanding of how a data archive could serve the data storage needs of researchers; or a wish to keep their data entirely under their own control. In both cases, advocacy and training allied with a strong mandate from the University could break down negative attitudes. As one respondent commented:

“If using a University data repository isn’t made compulsory it’s unlikely people will do it.”

³ Economic & Social Data Service (ESDS), a national data archiving and dissemination service, jointly funded by ESRC and Jisc

Overall, these results demonstrate an appetite and a need for advocacy and training to raise researchers' levels of awareness and skills in data management. There is a clear expectation by researchers that while data management is primarily their responsibility, the University has a definite role to play here, by providing support and training for researchers alongside a data archive system.

"I believe the uni is responsible to put systems in place and owners should be responsible to use them."

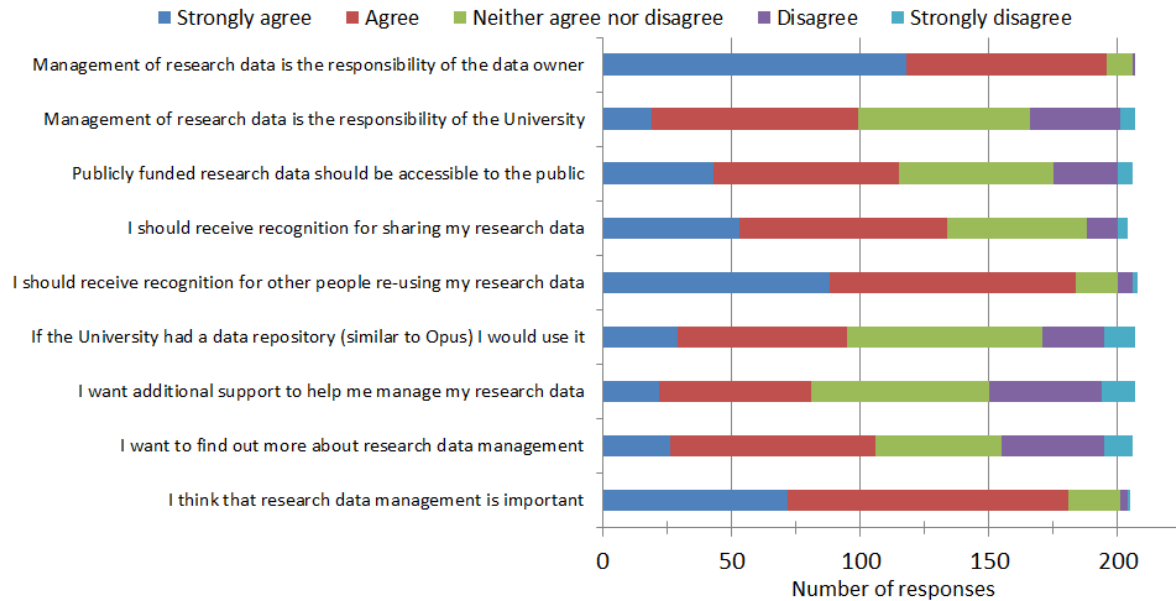


Figure 16: Researchers were asked whether they agreed or disagreed with a number of statements relating to research data management.

6. Case Study 1: Pilot Research Project in a Single University Department

The research

The researchers have carried out a pilot project collecting data from social media feeds and blogs written by a select group of individuals. This has been a proof of concept study for a funded project and the researchers are now putting together a funding application to a UK Research Council for a full 3-year project.

The research data

The social media data for the project consists of tweets and blog posts written by a select group of individuals. These are available over the internet and are harvested automatically. For Twitter stream harvesting the researchers are paying for the services of HootSuite, a company that specialises in the monitoring of social media streams, and then periodically archiving the Twitter data aggregated by HootSuite.

The downloaded data has been stored in MySQL databases at the University, on both centrally managed and departmental systems, and has been used to generate files for analysis with NVivo. The data must be preserved locally as, currently, Twitter streams become inaccessible after two weeks and blog posts can be altered or deleted by their authors. Estimated data requirements for this part of the project are small: probably a few gigabytes.

The researchers have also gathered information by interviewing participants in the study. These interviews have been recorded using digital audio recorders and the data transferred into audio transcription formats. The researchers estimate that about 20GB will be needed to store these data. The interview analysis requires the researchers to both transcribe dialogue and make additional notes on interviewee behaviour.

The full research project, if funded, would also require video recordings of participants using their current preferred social media client applications whilst at work. High definition video is preferred for such recordings as it allows zooming to various aspects of the media interface or desk environment without the loss of information. In uncompressed format, such recordings would consume about 1TB of storage per hour and approximately 30 hours of recordings are anticipated during the planned project. Current practice is to archive interviews on DV⁴ tapes that are stored in a locked cupboard in the researchers' office, while the transcripts and notes are kept in Word documents on resilient departmental storage systems.

⁴ A common format for storing digital video used by many video cameras

Research data management

The researchers are aware of the existence of funding council requirements for long-term archival of data through DCC guidance, brought to their attention by the Research360 research data survey. However, they are unsure of what this actually entails. Given the sensitive nature of the data involved, the researchers feel that a data management plan may help encourage participants to be more open than they might otherwise, as they can fully understand how their data will be protected.

As this was an internally-funded pilot project to test the data collection methods for a potential large funded project, it qualified for 1TB of centrally managed research storage. However, the initial internally-funded element of the pilot project terminated in March 2013 and additional exploratory work that was inspired by the pilot does not qualify for centrally managed storage. This highlights a need for provision of resource to support pilot projects, particularly if the data is vital to securing funding for larger projects.

Data from the pilot project is currently stored on the researchers' departmental workstations and servers. These are managed within the department and operated independently of the central computer services, raising important questions about how any systems and infrastructure implemented by BUCS will be regarded and adopted by departmental users.

The uncompressed video from interviews is expected to require around 30TB of storage. This could be improved using compression, but the scope for this is limited. The videos are an important record of interviewee behaviour and must include details from their computer screens and working environment. Scan converters or screen capture software do not record important details that indicate participant attention, such as gesturing or head and eye movements. Too high a level of compression would result in the loss of these essential details, rendering the video substantially less useful. The researchers currently do not know what level of compression will be suitable to retain sufficient resolution, whilst also reducing file storage requirements.

Audio and video interviews are covered by the Data Protection Act. Consent forms were used to record participants' agreement to be interviewed and, optionally, to be audio recorded. However, standard consent forms used for this type of research do not currently allow for long-term retention or publication of the data they cover and researchers felt that this should be included for future studies. Since, in theory, tweets and blog posts contain personally identifiable information, they may be also be covered by the Data Protection Act. However, when a tweet is posted the information it contains is made public. The legal implications of this for collection, storage, analysis, publication and archival of the data are currently unclear.

7. Case Study 2: Cross-Faculty Research Facility

The research

The research consists of two cross-department facilities that both bring together analytical instruments from across a Faculty to provide analysis services to researchers. In both cases researchers will either run their own samples or they will submit a sample to one of the Instrument Specialists within the facility, who will perform the requested analysis and return the results to the researcher in some form.

The research data

The type and amount of data varies widely depending on which instrument is used, from a single numerical value read off an LCD display, to several gigabytes. Most of the non-trivial data is stored in file-formats proprietary to the instrument vendor; even instruments of the same type will have vendor- and model-specific functions whose output must be recorded.

The majority of output files are large, so it is often impractical for researchers to take them away and take responsibility for the data. In addition, they must usually be analysed using proprietary software, for which only a handful of licenses may be available, so it is not often useful to take away the raw data in any case. Researchers will usually, therefore, use a dedicated PC near the instrument to perform their data processing and take away only the visualisations they need.

Estimates of the combined output of the facilities suggest that they will generate 4TB of data per year.

Research data management

The Instrument Specialists think that it is important to archive the raw data in the long term: it is common for researchers to come back and request data originally produced several years ago, and providing this archive is an accepted (and expected) part of the role of the analytical service. However, the Instrument Specialists currently feel they are coping rather than highly skilled in the area of research data management and they will therefore require additional training in research data management.

Because it is not practical or useful for researchers to take away their raw data, the Instrument Specialists have become de facto curators and archivists. However, a current draft of the research data management policy places responsibility for maintenance of data with the researchers and makes no mention of these facility staff. The role of facility staff in managing data needs to be addressed in the policy or associated procedure guide.

Although the facilities serve a large number of projects, they are not funded by discrete projects with associated cost codes. They are therefore not currently eligible for storage space under BUCS's current policy of 1TB per funded project.

When the facility was first established data storage was inconsistent, with data originally being stored on a combination of acquisition PCs, CDs, DVDs, current and defunct PCs and the X-drive. Backup and archival practice was equally varied. A total of 4TB of central storage had been made available to the combined facilities of the Faculty. Although this was

sufficient to archive currently existing data in a robust manner, it did not provide for the projected 4TB/year growth.

40TB of centrally managed storage has since been made available to the facility and existing data is being transferred from existing storage to this resource. However, this is proving to be a slow and technically difficult process: some acquisition PCs are not networked, requiring the necessary technical infrastructure to be put in place to enable data transfer. Facility staff are not able to dedicate the substantial time required to ensure accurate data transfer for each of the existing data storage points.

The case study highlights two important data management requirements for the future. Firstly, it will be necessary to identify resource to support data storage for other facilities within the University, possibly through a slight increase to the directly allocated or indirect charges for use of the facilities. Secondly, technical solutions are required to support the long-term backup and archive of data created by these facilities. Ideally, this should include automated back-up of data to centralised storage facilities. Additionally, a 'safe copy' of the data should be archived indefinitely whilst simultaneously ensuring that an accessible copy is available to researchers so that they can process their own data.

8. Case Study 3: Sharing Personal Data with UK and International Collaborators

The research

The research is a three-year multi-disciplinary project funded by a UK research funding council. The project is a collaboration between the University of Bath, five other UK universities, one of whom is leading the project, and a research institute in the USA. The project aims to combine a variety of data types involving a range of personal information to develop a large model.

The researchers are experienced in working with survey data and information covered by the Data Protection Act 1998. They have previously shared reports, but to date have not yet shared raw or processed research data outside of their research group.

The research data

A variety of data types will be generated during the project, including online surveys, biometric data, image recognition algorithms, interviews with key potential users of the model, and information gathered from social media. Not all of these data will be collected at the University of Bath, although there is likely to be some alignment of local data with that produced by other research partners. Early in the second year of the project, the researchers from all UK based partner institutions plan to collect a full dataset of all types of research data on a larger number of participants. This data will comprise one of the main datasets for the project.

Data relating to or gathered from social media will typically take several forms:

- In some situations participants will use simulated software to develop profiles, ensuring that personal data are not placed in third party environment and are not passed on to social media providers. Participants will be informed of this at the end of the exercise.
- Where real, existing profiles are used, frequency data such as the presence or absence of a particular piece of personal information will be the primary focus rather than content directly attributable to individuals. For this type of data collection, consent will be requested from website administrators. Data collection from websites is usually permitted within their terms of use, although there may be restrictions on the quantity of data that can be collected. Where content is attributable to individuals, data protection rules apply to ensure that it is processed fairly. For example, this may involve making sure that individuals know when personal data is being held about them and for what purpose.
- Two of the collaborating institutions intend to re-use existing datasets. One of these datasets was originally collected by the partner institution. The other dataset is being obtained from a third party within the EU.

The majority of the data will be in digital formats, although the researchers do not yet have a clear idea as to the size of the eventual dataset. Whilst survey data may be relatively small, biometric data and images used for recognition algorithms are likely to be substantially larger. Based on the researcher group's previous work, the physical storage space required

for non-digital data, typically paper-based consent forms, will be small, comprising only a couple of hundred sheets of A4 paper per study.

Research data management

There was no requirement to submit a formal data management plan with the funding application. Issues that would normally be covered by such a plan, including data storage, security and access, backup and publication are therefore being considered as and when the need arises. Currently, the project is formulating guidelines for requests under the Freedom of Information (FOI) Act 2000.

At each institute, project members are managing data under the legal rules applicable to data controllers as set out by the EU Data Protection Directive (95/46 EC) and the Data Protection Act 1998. 'Data controllers' are those who (either alone or jointly or in common with other persons) determine the purposes for which and the manner in which personal data are, or are to be, processed. A data controller's data protection strategy must include proactive data protection controls, which enable the ability to supervise and manage how personal data is kept in a way that is sufficiently secure. What amounts to sufficient protection of digital data depends upon their type and the risks to individuals if inadvertently disclosed but would usually include up-to-date technical measures, such as encryption.

In instances of collaboration between institutes, consent has been sought from the participant regarding the sharing of data with particular institutes, and all data are fully anonymised. Where an existing dataset has been provided by a third party, the terms for use and reuse of the dataset are restricted to the licence agreements made with the third party.

In accordance with data protection rules, consent forms (freely given and informed agreement by a data-subject to the collection and processing of their personal data) are required to be completed, without which the personal data generated by the project cannot be used. These also ensure that individuals are aware of the purposes for which their data is being processed, that their data shall not be kept for longer than is necessary for those purposes, and any third party disclosure policy.

Where online surveys are used, participants' consent is given via a tick box, which is therefore retained in digital format. Where paper-based consent forms are used, these are secured, restricted via a number of mechanisms to a limited number of individuals and retained as a single copy.

Importantly, data is being generated and subsequently analysed across two different jurisdictions with different legal rules – the EU and the USA. This affects how personal data is shared with the transatlantic partner institution in the research group.

Under EU law, special care must be given to the transfer of personal data collected within the EU to recipients outside the EU. Data protection rules require that personal data must be appropriately safeguarded so as to ensure an adequate level of protection when transferred to a non-EU country. These legislative considerations impact upon the use of data sharing technologies such as cloud storage, which are typically hosted in the USA.

The project has therefore been advised on relevant legal issues by a legal group based at one of the UK partner universities, including consideration of a number of secure data sharing options. One option included one UK partner institution setting up a networked drive, to which all participating researchers would have access. An alternative method involved

data being physically transferred between the partner institutions on a portable external hard drive. This hard drive would be encrypted with a limited set of encryption keys made available to key researchers in the project.

Due to the sensitive environment in which some interviewees work, interviews with potential users of the model will be manually recorded by researchers, enabling the scribe to anonymise information as it is captured. Interviewees will then be able to approve the transcripts prior to termination of the interviews. By using this approach, interview transcripts will be unclassified from the outset, thus eliminating the need to provide highly secure storage for the data or apply additional anonymisation techniques.

The researchers have begun to discuss plans for what will happen to the research data at the end of the project. Where it will not be possible to fully anonymise some of the biometric and image data, consent forms specify whether participants agree to their personal data being used for research and academic publication. If these data are to be made available to fulfil funder policy requirements, it would be under restricted conditions to specific individuals.

For other types of data, use is also constrained by conditions set out in consent forms, which limit the potential for re-use. However, consent forms do provide for long-term use (e.g. 3 years) of the research data after the end of the current research project, for example enabling re-use by the research group in any follow-up study. Licence agreements are also currently being prepared to enable third party research institutes to apply for access to parts of databases for limited periods of time under certain conditions.

9. Management View

Views and opinions on research data management were canvassed from 13 nominated senior stakeholders, representing a cross section of the main faculties and professional service departments within the University. Views were gathered via a series of semi-structured interviews, conducted by an external consultant, Neil Beagrie of Charles Beagrie Ltd, with the aim of informing development of a Business Case to support investment in research data management.

This section contains an anonymised synthesis of the interviews, together with recommendations and suggestions for a research data management strategy. Although a number of common themes emerged, not all of them were mentioned by every interviewee and there was some diversity in responses as well as points of commonality.

Institutional responsibility

It was recommended that there should be a permanent role responsible both for defining the services for the institutional data management framework and for on-going data management as projects complete and the researchers and supervisors move on. It was also suggested that a team take responsibility for defining and providing training in metadata and the use of additional data management tools. Some interviewees thought that while academics will be expected to add metadata themselves, a central monitoring role would be required to ensure that standards do not vary too widely between and within departments, as this would reduce future usefulness and accessibility.

It was recognised that it will be necessary to decide in which department this role/team should be positioned in. There was agreement that although BUCS can provide data storage facilities, they cannot define the framework or oversee the activities of individual academics in setting the parameters for retention periods etc. Data management is relevant to the Research and Development Office and they have the most extensive contacts with researchers on grant proposals and support. Research data management extends the traditional library role in information management so data management should involve part of the library/archivist remit. However, since library/archive professionals may not have sufficient expertise in file formats or data structures, preservation and storage, it was felt that they will need to work closely with IT colleagues. Some interviewees also thought that BUCS should advise on projects where secure sharing of sensitive data with selected external partners is required.

It was felt that a new research data management team should review all new research grant applications to ensure they request for appropriate funding to ensure the long term preservation of different research data types and that BUCS should assist in determining these costs. This was seen as essential for funding long-term data curation and access.

Some interviewees thought that initial effort of data management service provision should be focused on departments that had already perceived a need and associated benefits of improved data management.

Institutional Policies and Practice

One interviewee noted that with regard to an institutional data management framework, a distinction should be made between *active* and *archived* data. For example, data sharing policies might be defined so that access is restricted to a research team during the active phase, but access could be open for archived data. Several interviewees felt that retention and review periods should be considered and they questioned whether the new IT strategy to keep active research data for three years and archived research data for ten years would be appropriate in every case. It was suggested that transition arrangements would need to be agreed, such as embargo periods for either three or five years after the active phase.

For non-sensitive research data, it was suggested that the ultimate aspiration should be for researchers to click on a link in a published article to access the underlying data. For sensitive research data, the suggestion was that it may be appropriate for Data Protection compliance to define specific rules as to who can use the data for what purposes. It was also thought that consideration should also be given to anonymisation and for how non-specialists should be supported in handling requests for access to sensitive data.

A few interviewees also noted a need to agree levels of granularity for metadata and which fields and standards should be used. Similarly, levels of sensitivity and security require definition, as do intellectual property rights agreed with commercial partners.

It was recommended that institutional data management planning should seek to transfer current research data to storage managed by BUCS. Further, these plans should aim to also transfer legacy data to this storage and that data should be digitised if not already electronic.

Data storage and security

A common view amongst interviewees was that back-up and security of data were best handled professionally. Centralised storage managed by BUCS already offers high levels of resilience and appropriate security mechanisms for sensitive data. Many interviewees therefore agreed that this storage should be recommended and that the service should be reviewed regularly relative to new developments in the field. The interviews demonstrated that there is awareness of take-up of BUCS storage, but not of the expertise and services offered with respect to research data. Several interviewees' thought that a combination of academic secrecy and misplaced concern about the security of sensitive data on central servers had contributed to the perpetuation of suboptimal data storage practices both for live and legacy data. It was also thought that additional tools should be developed to enable academics to deposit data and supply associated metadata; to monitor data quality and variation; to alert those supporting research data management when intervention is required; and to identify stored data of interest to future researchers.

Open Access to Sensitive Data

The subject of data sharing raised concerns about handling and sharing information deemed to be sensitive from a commercial, personal, or national security perspective. Although several interviewees acknowledged that across the University there is good awareness of ethics issues and the University is a member of the UK Research Integrity Office (UKRIO), they were also concerned about the sensitivity of personal data in research, such as real

time GPS coordinates that could identify the location of a particular person, or projects that might touch on national security issues.

It was suggested that in order to satisfy commercial partners, data could be hosted by a trusted intermediary, but the academic PI cannot evade responsibility for responding to Freedom of Information requests, even if the data is not stored on University servers.

Provision of data management and metadata training

It was recommended that data management training be embedded in the induction and probationary processes for all new faculty staff and PhD research students. In addition, it was recommended that there should be a programme of refresher training or professional development for established staff. Some interviewees thought that this data management training should be compulsory for everyone up-to and including Associate Deans.

It was suggested that data management training should cover research data and its importance to the institution; the role of metadata in preservation and resource discovery; practical experience in assigning metadata; understanding risk and resilience in data security; best practice in encryption, back-up, storage and data sharing, including from mobile devices.

Cultural Change

There was recognition that there needed to be a major shift in perspective to recognise research data as a valuable institutional resource. This was vital to demonstrating research impact and therefore directly affecting future financial resourcing. It was agreed across a wide range of interviewees that awareness of data management needed to be raised in order to bring about the necessary cultural change in attitudes. Funder requirements for open access and research data management plans and likely legal requirements for FOI applying to research data were seen as key drivers for this. It was suggested that the direct benefits in terms of prestige, citations and facilitation of research-led teaching should be emphasised in order to encourage individual academics to invest time and effort in data management, change their data practices and relinquish perceived data ownership.

Diversity in data management practices

Many interviewees recognised that between different disciplines there is considerable diversity in data types, current practice and readiness for research data management. Science disciplines tend to produce complete datasets, whereas the Social Sciences may be supplementing, enhancing and repurposing datasets gradually over time. Departments such as Psychology have experience and understanding of data sharing and ethics issues, while in Engineering data is kept for research integrity only, not generally suitable for reuse after the active phase and more likely to have commercial restrictions.

It was recognised that while any centralised data management service would need to take account of disciplinary variations, different standards and approaches between the disciplines will be an obstacle to the establishment of a true multi-disciplinary repository.

10. Recommendations

The recommendations presented here are the result of survey responses including free-text comments, a small number of case studies and a synthesis report based on interviews with senior stakeholders. These draw a picture of the current environment for research data management practices at the University of Bath, and give a flavour of the perceptions on the topic. Whilst not all questions deal specifically with industrial and commercial collaborations, this theme is drawn out repeatedly as a point of interaction and activity amongst University researchers, demonstrating a need for clarification of University policy, procedures, available resources and training on this topic.

The survey explored the following themes:

- Data ownership
- Data management planning
- Current data storage
- Data loss
- Data sharing
- Data archive
- Perceptions

Based on the responses to these themes, the following recommendations are suggested:

1. Infrastructure and services should be further developed to support researchers at Bath, particularly with reference to meeting expanding data storage needs, funder requirements for access to data and the requirements of the 'University of Bath Roadmap for EPSRC'. As much of the research undertaken at the University of Bath is collaborative in nature, this infrastructure must include suitable tools to facilitate the sharing of data during the active phase of the project.
2. Training and resources should be enhanced to develop graduates with good data management skills, including an understanding of the issues arising from collaboration and future sharing of data. Good quality training resources are needed both within structured training schedules, and at the point of need. Guidance on current University provision for data management and also on the appropriateness of third-party services must be made readily accessible.
3. Clarification on data ownership issues, particularly when negotiating with commercial partners, should be provided. This will help reduce confusion over what can or can't be done with the data throughout the research lifecycle. This clarification should include both a standard University policy on data management and also signposting for help in case of issues that fall outside of regular policy or procedures, for example when non-disclosure agreements run counter to research council or other funder agreements.
4. Data management plans should be used more widely to focus processes for meeting funding or industry requirements. Templates and supporting guidance should be made readily accessible for use both during the grant application process and by research postgraduate students.

5. An outline of the benefits of good data management, both to the University community and to external partners such as industry and commerce, would provide clarification and support for the activity for all stakeholders.
6. A common thread amongst all themes and responses was the desire for institutional support throughout the research lifecycle. Resource should be invested not only in technical infrastructure but also in professional support. The latter would provide infrastructure development, training researchers in good data management practices, and ensuring compliance with funder and industry requirements.

11. References

- Ball, A., 2010. *Review of the State of the Art of the Digital Curation of Research Data*. University of Bath. (ERIM Project Document; erim1rep091103ab10). Available from : <http://opus.bath.ac.uk/19022/> [Accessed 9 July 2013].
- Ball, A., 2012. *Tools for research data management*. Bath, UK: University of Bath. Available from: <http://opus.bath.ac.uk/29189/> [Accessed 9 July 2013].
- Beagrie, N. and Pink, C., 2012. *Benefits from Research Data Management in Universities for Industry and Not-for-Profit Research Partners*. Charles Beagrie Ltd and University of Bath. Available from: <http://opus.bath.ac.uk/32509/> [Accessed 9 July 2013].
- Digital Curation Centre. 2011. *Collaborative Assessment of Research Data Infrastructure and Objectives (CARDIO) Benchmarking tool* [online]. Available from: <http://www.dcc.ac.uk/projects/cardio> [Accessed 9 July 2013].
- Jones, K., 2011. *Assessing institutional data storage and management using the Data Asset Framework (DAF) methodology at the University of Bath*. Bath: University of Bath. Available from: <http://opus.bath.ac.uk/24960/> [Accessed 9 July 2013].
- Lyon, L., 2007. *Dealing with Data: Roles, Rights, Responsibilities and Relationships. Consultancy Report*. UKOLN. Available from: <http://opus.bath.ac.uk/412/> [Accessed 9 July 2013].
- Lyon, L. and Pink, C., 2012. *University of Bath Roadmap for EPSRC : Compliance with Research Data Management Expectations*. Bath: University of Bath. Available from: <http://opus.bath.ac.uk/31279/> [Accessed 9 July 2013].
- Pink, C., 2012. Working with Commercial Partners (Presentation). In: *British Library DataCite Workshop - Managing and citing sensitive data*, 29 Oct 2012, London. Available from: <http://opus.bath.ac.uk/32291/> [Accessed 9 July 2013].

12. Appendix 1: Survey Questions

University of Bath Research Data Survey

Introduction

Welcome to the University of Bath Research Data Survey

Thank you for taking part in this survey, which asks about the source materials (data etc.) which is the raw material for your research. Your answers will assist the University in providing services and support to help you keep your research data and other source materials safe, and make those materials accessible to collaborators or the public according to your needs and the requirements of your funder.

Throughout the survey, we use the generic term *research data* to refer to **any** source information on which research is based, including but not limited to:

- Results of experiments or simulations
- Statistics and measurements
- Models and software
- Observations e.g. fieldwork
- Survey results -- print or online
- Interview recordings and transcripts, and coding applied to these
- Images, from cameras and scientific equipment
- Textual source materials and annotations

This survey is part of the Research360 project, for which the University has received external funding from Jisc. The project is founded on the principle that research data is a key asset for the University and its researchers, and aims to develop technical and human infrastructure with accompanying policy and guidance to enable researchers to make the best possible use of that asset.

How to answer this survey

Some questions in this survey should be answered from the perspective of a particular research project. Please identify a single project that is typical of your research and refer specifically to that project in any sections which ask about "your chosen research project".

You have the choice to give your feedback anonymously or to provide your username. Any responses will be made anonymous for analysis. Survey data may be collated and summarized for external publication.

We may wish to contact you for a more detailed follow-up interview to better understand your needs. If you are happy for us to contact you in this way, please enter your username where requested. If you do choose to supply your username, you will be entered into a prize draw to win an **Amazon Kindle 3G**.

Section 1: About you & your research

The information you provide in this section will enable us to assess how research management perceptions and requirements differ between different disciplines, stages of research and project types.

1. What department or school do you belong to? [Drop down menu]

- Architecture and Civil Engineering
- Biology & Biochemistry
- Chemical Engineering

- Chemistry
- Computer Science
- Economics
- Education
- Electronic & Electrical Engineering
- Department for Health
- Management
- Mathematical Science
- Mechanical Engineering
- Natural Sciences
- Pharmacy & Pharmacology
- Physics
- Politics, Languages & International Studies
- Psychology
- Social & Policy Sciences
- Other [Free text box]

2. Which role best describes you?

- Taught postgraduate student
- Integrated PhD student
- Research postgraduate student
- Research Officer (post doc. Research associate)
- Principle Investigator (PI)
- Other academic staff
- Technician
- Other (please specify) [Free text box]

3. How many projects are you currently involved with?

- 1
- 2
- 3-4
- 5 or more

4. Who funds your chosen research project? (select all that apply)

- University of Bath
- A single UK funding body e.g. research council
- A single international funding body e.g. NIH, European Union
- Multiple funding bodies
- Industrial/commercial sponsor
- Internal funding
- I don't know
- Other (please specify) [Free text box]

5. Who are you working with on your chosen research project? (select all that apply)

- I work completely independently
- I work only with my supervisor(s)
- collaborators within my department
- collaborators in other departments within the University
- collaborators at other universities
- collaborators in commercial industry
- collaborators in the public sector
- collaborators within the UK
- collaborators internationally
- I don't know

- Other (please specify) [Free text box]
- 6. What computer/device do you use to generate, access and/or use your data? (Optional) (select all that apply)**
- A Windows PC
 - A Linux (or other Unix) PC
 - A Mac
 - A standard mobile phone
 - A smartphone (e.g. iPhone, Android)
 - A tablet computer (e.g. iPad)
 - Other (please specify) [Free text box]

Section 2: Data ownership

This section will help us to understand how ownership of research data is defined across the University.

*Please answer the following questions relating to **your chosen research project***

- 7. What are the most important source of your research data? (Optional) (select all that apply)**
- Your own research data (generated or collected by you)
 - Data shared by a colleague at Bath
 - Data shared by a colleague in another university
 - Data you obtained from supplementary data associated with a publication
 - Data you obtained from a national or international data repository
 - Data provided by a commercial or public sector research partner
 - Other (please specify) [Free text box]
- 8. Management of research data involves ensuring its quality and security of storage, describing, archiving and publishing it. Who is ultimately responsible for these activities? (Optional) (select all that apply)**
- You
 - Your project supervisor/Principle Investigator
 - The owner of the Intellectual Property
 - A designated data steward
 - The University
 - I don't know
 - Other (please specify) [Free text box]
- 9. If you have any other comments about the ownership of your research data that have not already been covered, please give details here (Optional) [Free text box]**

Section 3: Data management planning

This section is about how you plan to organise your data. Funding bodies increasingly require grant-holders to develop and implement Data Management Plans (DMPs).

Data management plans typically state what data will be created and how, and outline the plans for sharing and preservation, noting what is appropriate given the nature of the data and any restrictions that may need to be applied.

Reference: Digital Curation Centre, n.d. Data management plans [online]. Edinburgh: DCC. Available from: <http://www.dcc.ac.uk/resources/data-management-plans>. (Accessed 02/05/2012)

Please answer the following questions based on your general knowledge and experience

10. Have you ever produced a data management plan (see description above)? (Optional)

- Yes, I chose to do so
- Yes, I was required to do so
- No (Please go to Section 4: Data Security)

11. I was required to produce a data management plan by: (Optional) (select all that apply)

- University of Bath policy
- University of Bath Code of Good Practice in Research
- University of Bath guidelines
- Funder guidelines or requirements
- Publisher guidelines or requirements
- None of the above
- Other (please specify)

12. What type of template did you use to generate your data management plan? (Optional)

- I did not use a template
- Based on guidance provided by the funding body as part of the application form
- DMP Online
- DMP Tool
- DataTrain
- Copied from an existing data management plan from my project group
- Copied from an existing data management plan that I found elsewhere
- Other (please specify) [Free text box]

13. Did you refer to your data management plan after you had produced it? (Optional)

- Yes, but I didn't find it of any use
- Yes, and it assisted me in managing my data throughout the project
- No
- I don't know

14. If you have any other comments or questions about data management planning that have not already been covered, please give details here (Optional) [Free text box]

Section 4: Data security

This section refers to the nature of your data, and whether there are any special conditions that might apply to it which may impact on your ability to freely share it. For example, your research may have required approval by the ethics committee, or be subject to a contractual agreement drawn up between you and a collaborator.

Please answer the following questions based on your general knowledge and experience

15. Is your data...? (Optional) (select all that apply)

- Subject to a nondisclosure agreement
- Subject to the Data Protection Act
- Commercially exploitable or of other potential commercial value
- Potentially controversial (e.g. genetic modification, tobacco, animal based, climate change, human subjects requiring anonymisation)
- Confidential for other reasons (please give details under 'Other' below)
- Other (please specify) [Free text box]

16. If there are specific security requirements imposed on your research from outside the University (e.g. by collaborators or funders), please give details (Optional) [Free text box]
17. If you have any other comments or questions about data security that have not already been covered, please give details here (Optional) [Free text box]

Section 5: Data management

This section is about how you look after your research data during a project. This will include how you generate, store, organise and analyse your research data. We will be asking about how you work on your research data with any collaborators in the next section.

Please answer the following questions relating to your chosen research project

18. If you have any non-digital data, what format is it in? (Optional) (select all that apply)
- All of my data is in a digital format
 - Paper (text or image, e.g. lab notebook)
 - Analogue media e.g. tape
 - Model
 - Samples e.g. enzymes
 - Live samples e.g. fly, plant, rodent or bacterial strains
 - Other (please specify) [Free text box]
19. Which types of storage do you use to store your digital research data? (Optional) (select all that apply)
- The X:\ drive (shared research project filestore)
 - The H:\ drive (personal filestore)
 - My University of Bath computer's hard drive
 - Other University of Bath filestore (please tell us which under 'Other' below)
 - My home computer's hard drive
 - External hard drives
 - USB sticks
 - CD or DVDs
 - Dropbox or other online/cloud storage
 - My collaborators manage all of my research data
 - Other (please specify) [Free text box]
20. Are there any reasons why you would not use storage space managed by the University of Bath (e.g. the X:drive)? (Optional) (select all that apply)
- I didn't know it existed
 - I have concerns about security
 - I have concerns about resilience
 - I have concerns about cost
 - I have concerns about speed, connectivity or offline access
 - There is insufficient space available
 - Lack of version control
 - None of the above
 - Other (please specify) [Free text box]
21. Approximately how much of your research funding is spent on storage (e.g. external hard drives) for your digital research data per year? (Optional)
- Nothing
 - £50 or less
 - £50-£100

- £100-£300
- Over £300
- I don't know

22. Having stored your data, how have you described it so that it can be understood? (Optional)

- Based on a direct attribute of the data e.g. column names
- Using information in the filename
- Using information directly attached to the data (e.g. metadata)
- Using descriptions within an additional file stored with the data e.g. readme.txt
- I do not keep records of what my data is

23. I would be able to interpret my data meaningfully in 3 years' time (Optional)

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

24. Other people would be able to interpret my data well enough to re-use it (Optional)

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

Section 6: Data loss

This section will enable us to understand more about how research data is lost, so that we can work towards reducing the risk of this happening in the future.

Please answer the following questions based on your general experience

25. Have you ever lost any research data?

- No (Please go to question 26)
- Yes

a. If yes, how frequently have you lost research data? (Optional)

- Only once
- 1-3 times in total
- About once per year
- Approximately monthly
- More frequently than once per month
- I don't know
- Other (please specify) [Free text box]

b. If yes, how much data did you lose? (Optional)

- At most, a day's worth of work
- Approximately several days' worth of work
- Approximately a month's worth of work
- An entire research project
- Everything
- I don't know
- Other (please specify) [Free text box]

c. If yes, why did you lose the research data? (Optional) (select all that apply)

- A virus

- A hardware failure
- Physical damage e.g. fire, washing a USB drive
- Loss of non-digital data e.g. lab notebook
- Hardware theft
- Hacking
- File corruption
- Accidental deletion of the only existing copy
- Storage in an obsolete file format e.g. WordPerfect
- Storage on obsolete storage hardware e.g. floppy disk
- I don't know
- Other (please specify) [Free text box]

26. What processes do you have in place to avoid data loss? (Optional) (select all that apply)

- None – my data could be easily replaced if it were lost
- Utilisation of University of Bath managed storage e.g. X:drive
- Backed up on an external hard drive
- A second copy on my own computer
- Additional copies on a second computer
- An additional version on an alternative media type e.g. USB stick, CDs
- Use of cloud storage e.g. Dropbox, Amazon S3
- Scanned copies of non-digital paper media
- None of the above
- Other (please specify) [Free text box]

Section 7: Data sharing

This section is about how you work with other people on your data, during the life of your research project. The information you provide will help us to develop future infrastructure to meet your needs.

*Please answer the following questions relating to **your chosen research project***

27. How much of your data can you share with other people directly involved in your research project? (Optional)

- None of it (Please go to Question 31)
- Some of it, but specific elements withheld
- Only what is required by my collaborators
- All of it
- I don't know
- Other (please specify) [Free text box]

28. How do you provide others with access to your data? (Optional) (select all that apply)

- Email
- On paper
- DVD, USB stick, external hard drive or other physical digital media
- University shared drive e.g. X:drive
- Personal university web page, iSusLab or other university-provided service
- Personal non-university web page
- Dropbox or other sharing service provided by a third party
- Service provided by a collaborator
- I don't know
- Other (please specify) [Free text box]

29. What volume of data do you share with others? (Optional)

- Less than 10MB
- 10-100MB
- 100MB-1GB
- 1GB-1TB
- More than 1TB
- I don't know

30. If you have any restrictions on the sharing of data in this way due to the nature of your research or your collaborators, please give details here (Optional) [Free text box]

31. How do you track different versions of your files? (Optional) (select all that apply)

- I do not create different versions of my files (I re-save a single file as I update it)
- File- or folder-naming conventions (e.g. midterm-report_v5.doc)
- Automated backup software (e.g. Time Machine)
- A document management system (e.g. Microsoft SharePoint)
- Dedicated version control software (e.g. CVS, Subversion, Git, Mercurial)
- I don't know
- Other (please specify) [Free text box]

Section 8: Data publishing and archive

This section is about what you do (or will do) with your research data once you have completed the project. It will also give us an idea about where your completed research data is kept and whether you have been asked to share your data with people outside your research group. This will help us to understand what you do with your research data, so that we can plan for future data storage requirements.

Section A: Current projects

Please answer the following questions relating to a current research project. If you are not working on a current research project, please go to section B.

32. Are you planning to make your primary research data publicly available for download? (Optional)

- Yes, because I am required to do so
- Yes, because I would like to
- No, I don't want to
- No, I'm not allowed to
- I don't know
- Other (please specify) [Free text box]

33. How do you plan to make your primary research data publicly available for download? (Optional) (select all that apply)

- I don't plan to make my research data publicly available
- Deposited in an external discipline repository
- Deposited in my funder's repository
- As supplementary data accompanying publication of a scholarly article
- Published in a data journal
- On my personal website
- On a collaborator's website
- I don't know
- Other (please specify) [Free text box]

34. Which of these restrictions apply to the availability of your data? (Optional) (select all that apply)

- There are no restrictions
- Time embargo
- Anonymisation to remove subjects names
- Unable to make data available due to confidentiality agreement
- Unable to make data available to protect future commercialisation/patenting
- I don't know
- Other (please specify) [Free text box]

Section B: Completed projects

Please answer the following questions relating to completed research projects. If you have not completed any research projects, go to Section C.

35. Have you kept any research data from completed projects? (Optional)

- Yes
- No
- I don't know

36. How much data (in MB/GB/TB) do you capture in total for a typical research project? Tip: you can find this out by right-clicking the project folder and choosing "Properties..." (on Windows) or "Get Info" (on Mac) from the menu (Optional)

- Less than 10MB
- 10-100MB
- 100MB-1GB
- 1GB-1TB
- More than 1TB

37. How have you made your primary research data publicly available for download? (Optional) (select all that apply)

- I have not made any of my research data publicly available
- Deposited in an external discipline repository
- Deposited in my funder's repository
- As supplementary data accompanying publication of a scholarly article
- Published in a data journal
- On my personal website
- On a collaborator's website
- Other (please specify) [Free text box]

38. Has any of your publicly accessible research data subsequently been reused? (Optional) (select all that apply)

- Yes, I have re-used the data myself for a different research project
- Yes, it has been re-used by a different research group, but I did not receive credit as the original data creator
- Yes, it has been re-used by a different research group and I have received credit as the original data creator
- Not applicable as I have not made any research data available publicly
- No
- I don't know

39. Have you ever been asked for your unpublished research data? (Optional) (select all that apply)

- Yes – through an informal request and I provided the data
- Yes – through an informal request but I did not provide the data
- Yes – under Freedom of Information and I released the data

- Yes – under Freedom of Information but I did not have to release the data
- Yes – under Environmental Information Regulations and I released the data
- Yes – under Environmental Information Regulations but I did not have to release the data
- No
- I don't know
- Other (please specify) [Free text box]

40. If you are required by law to keep your data (e.g. consent forms, aeronautical records), please give details here. (Optional) [Free text box]

Section C: Non-digital data

Please complete this section if you have research data stored in paper (hard copy) form

41. How long do you need to keep your non-digital data? (Optional)

- Not at all
- Less than 5 years
- 5-10 years
- 10-20 years
- 20-30 years
- More than 30 years
- I don't know
- Other (please specify) [Free text box]

42. How much space does your non-digital research data take up? (Optional)

- Approximately a medium sized box
- Approximately a filing cabinet
- A cupboard full
- More than a cupboard full
- I don't know
- Other (please specify) [Free text box]

43. How are you currently storing this data? (Optional)

- In your office on campus
- In additional storage space on campus
- Using your personal space e.g. at home
- In third party storage space off campus e.g. with storage contractor
- Other (please specify) [Free text box]

44. If you pay for off campus storage, how much does this cost per year? (Optional) [Free text box]

Section 9: General Questions

The study and formalisation of research data management is a relatively new area. As we start to work on improving the infrastructure in place to enable you to do this more effectively at the University, it would help us to understand more about what your thoughts are.

45. Please tell us whether you agree or disagree with the following statements:

- Management of research data is the responsibility of the data owner
- Management of research data is the responsibility of the University
- Publicly funded research data should be accessible to the public
- I should receive recognition for sharing my research data
- I should receive recognition for other people re-using my research data

- f) **If the University had a data repository (similar to Opus) I would use it**
- g) **I want additional support to help me manage my research data**
- h) **I want to find out more about research data management**
- i) **I think that research data management is important**

For each question, provided answers are:

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

46. If you have any other comments, questions or requests about the University's research data support, please share them with us here.

We are particularly interested in research data issues arising from interaction with **industry and other external organisations (Optional) [Free text box]**

And finally

47. Results will be anonymised and no names will be published in any reports.

If you would like to be entered into a draw to win a **Kindle Touch 3G**, please enter your **University of Bath login (BUCS ID)** below. (Optional) [Free text box]

48. Would you be happy for us to contact you with some follow up questions so that we can learn more about how you manage your research data? (Optional)

- Yes
- No

Final Page

Thank you for taking the time to help us by completing this survey.