Shared Storage Requirements Project

Final Report

Document Information

Document Version:	1.0
Date:	12th February 2015
Author(s):	Merv Jones – Cambridge SMC Ltd
Produced For:	Bob Dowling – University Information Services (UIS)
Document Status:	First released version

Contents

1. Document Introduction 4
2. Project Approach 4
3. Funding of storage 5
4. Existing storage capacity 6
5. File sharing and synchronisation – current state 7
5.1. Dropbox 7
5.2. OwnCloud 7
5.3. Other cloud-based file-sharing solutions 8
5.4. CamTools 8
5.5. Sharing of larger data sets 9
5.6. Sharing of data with students 9
5.7. Use by administrators 10
5.8. User interest in a UIS Dropbox-type service 10
5.9. User requirements for a UIS Dropbox-type service 10
5.10. User types and suggested quotas 11
6. Archiving and long term data retention 13
6.1. Data Retention - Sponsor Requirements 13
6.2. Data Retention – Researcher Requirements 14
6.3. The current approach to archiving 15
6.4. Researchers concerns about archiving 15
6.5. Metadata 16
6.6. Future approach to archiving 16
7. Backup – current state 18
8. Backup – future requirements 19
9. Regulatory Compliance and Data Security 19
10. General Concerns 20
11. Manners of access to storage 21
12. Availability 21
13. Authentication and authorisation 22
14. Metrics 22
15. Performance 23
16. Related projects 23
17. Other services 24

18. Conclusions 24

Appendix A – List of participants 26

1. Document Introduction

This document is the output of a 3 month spell of intense activity performed on behalf of University Information Services (UIS) between November 2014 and February 2015 to understand the University's research user requirements for services based upon a shared storage infrastructure. The terms of reference for the project were set out in a separate document entitled "*DRAFT Storage requirements capture project (v1*)" dated August 2014 – extracts from which are included below:

"Executive summary

Resource is required to undertake a requirements capture for online storage in the University of Cambridge. This would be a "snowball" project where consultants would develop additional contacts from their interviews with contacts already on a list, with a goal to developing representative sampling from across the university's research groups. Specific data is required but additional notes from each meeting would be welcome.

Detailed project

The University expects to develop a number of online storage solutions for its researchers, teachers, students and administrators. It is recognised that no single solution can fit all needs so the UIS will undertake a detailed requirements capture project to determine what set of solutions most closely meet the majority of needs. To this end we require consultants to interview a large number of PIs ("principal investigators", the named holders of research grants) to determine their needs. We will start the consultants with a set of named contacts in various University institutions. This list will not be exhaustive and we would expect the consultants to generate other contacts from the people they interview. Over the course of the interviews the consultants would be required to maintain a balance over the six academic Schools of the University and also to engage with the Schools' own IT Coordinators.

It is likely that a number of the researchers will be unable to give direct answers to the questions. The interviewers will need the skills to interpret answers dressed in academic requirements into technical requirements. It is likely that the IT staff interviewed will not have these difficulties. While the interviews will be a principal contributor to the development of storage solutions for the collegiate university, it is essential that no specific expectations be raised."

2. Project Approach

There are two typical approaches to gathering user input in this type of project – either via a questionnaire and/or through interview. Both approaches clearly have their pros and cons, but owing to the diverse roles and requirements of those involved in the process and the need to be able to interpret individual responses *(and dig further as required)*, an interview approach was regarded by University Information Services (UIS) as being the most appropriate.

A list of circa 20 individuals was initially provided in order to set the discussions in motion. The vast majority of these initial interviewees were in an IT-oriented role *(typically IT Manager or Computer Officer),* and were asked as part of the discussions to propose some additional names for interview from a research perspective. These individuals were added to the schedule and subsequently interviewed in order to understand their specific end-user requirements.

Aside from two telephone interviews and another couple who preferred to meet at the Roger Needham Building, all discussions were conducted at the work place of the interviewee in order to minimise the impact on their working day. Although all six Schools of the University have been included in this process, it has not been possible to interview people from every single department – as this would have significantly increased both the project costs and overall duration. Nevertheless, though performing well in excess of 100 interviews over a 3 month period, a wide cross section of roles and departments has been included – which should prove to be representative of the wider University's needs.

Full details of participants and the areas they represent are listed in Appendix A

NOTE: For ease of reading, feedback from these interviews has been grouped into sections – which closely reflect the structure of the original terms of reference.

3. Funding of storage

The funding of storage is a big issue owing to the ever-increasing amount of data being produced and is an area where many researchers have not really helped themselves. The view from most of the IT interviewees is that researchers have not been particularly good historically at ensuring that storage costs are included on their grant applications. Whilst adequate costs for computing hardware and software – and even services such as High Performance Computing (HPCS) are typically being included on the grant application, storage has been regarded by most as a free 'utility' *(in the same way as they see heating, water and electricity)* and as such is not being considered as an additional cost of performing research. This omission puts ever-increasing pressure on the IT departments to maximise the storage space that they have available, and invariably ends up in quotas being introduced in order to restrict the capacity available to users. The introduction of quotas can sometimes drive undesired behaviour however – and increases the risk of data loss with many PIs opting to buy additional 'cheap USB storage' rather than investing in what they regard as 'expensive central storage' which is regularly backed up. *(One researcher stated that his IT department would charge him £1,000 per TB for additional central storage capacity, whereas he could go to PC World and buy a 4TB USB Drive for little over £100.)* Although not exactly comparing apples with apples, this price differential is huge - and is something that many PIs simply cannot resist when money is tight owing to the fact they have omitted to factor in sufficient storage costs for their research projects.

Some IT interviewees stated that things were beginning to improve slightly however, and that PIs were now increasingly seeking their advice and guidance on storage requirements and costs whilst putting together their grant applications. As a typical research grant is between 3 and 5 years in duration though, it might be some time before the historic omission of storage costs has flushed through and any recent changes in behaviour start to make a noticeable difference. Additional steps such as further raising user awareness and adding storage costs to a checklist when internally reviewing a grant application *(a process recently introduced in one department)* could also go a long way to making sure that we get sponsors to fund as much of the ongoing costs of storage provision as possible.

Once some central storage options are made available by UIS, the introduction of a simple pricelist/menu would be welcomed by researchers as this would allow them to very easily calculate the cost of their storage needs and ensure that this is factored into future grant applications.

Whilst there is a real interest in the potential of being able to buy storage as a service from a central provider, there is also concern that the costs of doing so might not be competitive. This especially relates to those individuals/groups that currently have their own local storage provision – much of which is on very inexpensive USB drives. The typical researcher approach to storage right now is very much a case of *"I only have this much money available – how far can I make it stretch?"* One of the biggest challenges we face is in changing this mind-set to a more proactive approach of *"this is how much storage my project is going to need – how much is it going to cost – and who is going to pay for it?"*

4. Existing storage capacity

As one would expect, the current storage capacity varies tremendously from department to department and from group to group. At one end of the scale, we have a Faculty in the School of Arts and Humanities with 30TB of storage capacity – with just 5TB currently being used. They estimate that they have enough space for at least the next 5 years.

At the other extreme, CRUK have 2.2PB of storage *(of which 1.8PB is currently used)* and are already looking at adding more capacity in 2015.

IT interviewees were asked how long their current storage would last before they needed to add more capacity, and aside from a couple of exceptions who had fairly urgent requirements, the typical range was between 18 months and 3 years – with a minority believing they could manage for another 5 years or more. IT users reported that they were more than happy to consider buying future storage from a central provider rather than continue to add further capacity at a local level – so long as it was more cost effective to do so and that access speeds were adequate.

What complicates matters somewhat is the number of research groups who have their own local storage which is not managed centrally by their Computer Officer/IT Dept and is therefore largely 'invisible' to them. In some cases, this is because there is no central departmental storage provision *(e.g. Material Science & Metallurgy, Chemical Engineering & Biotechnology)* and in others it is because individual research groups have opted to add more storage locally to complement what the department already provides them with. Some groups have their own local servers or have purchased their own Network Attached Storage (NAS) boxes, but there are also an alarming amount of USB hard drives in use – which are not backed up in any way and pose a significant risk of data loss. Although the risk of using such methods is understood and accepted, the low cost per TB of USB storage is proving to be irresistible in many cases.

Adding up the total storage capacity of those interviewed *(including both central storage and any additional local storage held by the researchers)* gives us in the order of 15PB. Owing to the amount of 'invisible' storage and the fact that not every single department and researcher was visited, the overall storage capacity of the University will be significantly higher than this however, so any solutions that UIS introduce will need to be readily extensible.

5. File sharing and synchronisation – current state

As initially suspected, the use of cloud-based file sharing and synchronisation solutions such as Dropbox, OwnCloud, and GoogleDrive etc. is widespread across the research community. Aside from just a couple of exceptions, every single interviewee used one or more of these solutions to a greater or lesser degree, with the vast majority *(>85%)* using Dropbox.

Even those areas where the use of such solutions might not be appropriate *(e.g. owing to data security and ownership concerns)* made use of them – despite frequent reminders and recommendations from their IT staff that they should not be utilised. Where this was the case however, researchers were very keen to state that they restricted their use and took great care to ensure it was limited only to such files that were in no way sensitive or confidential.

5.1. Dropbox

Of those researchers currently using Dropbox, a good proportion *(approx. 30%)* were 'heavy users' and had upgraded to the 'Dropbox Pro' edition – offering them up to 1TB of storage at a cost of ~£8 per month. This was deemed to be an acceptable overhead for the capacity and features provided, with many stating they were happy to fund the cost themselves and not attempt to recoup it in any way. A couple of other 'heavy users' were using GoogleDrive instead as they favoured its in-application editing capability – a feature which Dropbox doesn't support.

Remaining Dropbox users subscribed to the 'Dropbox Basic' free option, with many having more capacity available than the basic 2GB allowance through recommending it to friends/colleagues and also participating in the 'Dropbox Space Race' back in 2012. A free quota in excess of 10GB was not uncommon amongst this group. Dependent on how they used the solution, such users sometimes had to delete items when they ran out of space, but were loath to pay for an upgrade to 'Dropbox Pro' and were perfectly able to cope through being selective about what was being added to their Dropbox account.

In terms of features and usability, all Dropbox users were highly complementary about the solution – not only was it regarded as being very easy to use, but it was highly reliable and to quote one interviewee "just worked!".

NOTE: Any increased free quota resulting from the 'Dropbox Space Race' was initially due to expire on 15th January 2015, but has very recently been extended to 4th March 2015. Link to relevant terms & condition provided here. After this date, users will automatically revert back to the basic account levels again – a fact that a great many were not aware of until it was mentioned to them.

5.2. OwnCloud

A handful of departments *(CIMR, Haematology, Stem Cell Institute, Chemistry, Physics, Engineering, MRC and Geography)* had either recently implemented or were in the process of piloting their own local solutions based on OwnCloud. Uptake of such solutions was reported to be fairly slow – with a lack of marketing / communication effort and difficulty in 'wrestling' users away from their existing solutions believed to be the main reasons for this. IT depts. also confirmed that their launches had

intentionally been very low key affairs - to ensure that uptake was gradual and any teething issues could be resolved without too much adverse user impact.

One department *(Gurdon Institute)* had previously embarked on a small pilot of OwnCloud in early 2014 but had parked it after a short time and gone back to using DropBox again owing to issues with slow sync speeds when transferring folders containing large numbers of small files.

A small pilot of OwnCloud with 8 to 10 Colleges *(driven by the College IT Management Group – CITMG)* was also reportedly getting underway – although it was too early to obtain any feedback as to the success of this activity.

A few other departments were also considering their own local OwnCloud implementation, but were holding back right now to see what came out of this review – in the hope that the University would deliver a central solution instead.

Although the use of OwnCloud across the University appears to be fairly limited right now, it is understood to offer comparable features and user experiences to that of Dropbox – with the added benefits of being able to host it on our own servers and also having in-application editing capabilities.

5.3. Other cloud-based file-sharing solutions

A small number of the researchers interviewed used alternative solutions for file sharing and synchronisation – either instead of or as well as Dropbox / OwnCloud. *(A couple admitted to subscribing to multiple free versions of such solutions in order to maximise the amount of free cloud-based storage available to them).*

These alternatives included Box, OneDrive, iCloud, Oxygen Cloud, Spider Oak, Wetransfer, and Google Drive. Aside from the ability of GoogleDrive to offer in-application document editing *(which OwnCloud offers but Dropbox doesn't),* none of these appears to offer anything different and is purely down to a matter of personal choice in the absence of any centrally-provided University solution.

5.4. CamTools

CamTools was also mentioned by a small number of interviewees *(~8)* as being a solution that they used in order to share data. Usage patterns were varied. One PI used it every day as his main method of collaboration and communication – not only across his research group but also with trusted external parties. Others reported that they used it now and again as a means of sharing selected information only.

A couple of these individuals had heard on the grapevine that CamTools was going to be retired in the not too distant future and were keen to understand exactly what was going to replace it. The accuracy of such rumours has not been verified and there is nothing on the CamTools home page to suggest this is the case – so there is a chance that their concerns might be unfounded.

Although the user experience of CamTools was not particularly well liked *(the words 'clunky' and 'outdated' were both used)*, it was regarded as a useful tool nevertheless. If it is to be retired in the near future we need to make sure that our intentions are well communicated

5.5. Sharing of larger data sets

Whilst solutions such as Dropbox and OwnCloud are perfect for sharing individual files and smaller data sets, they don't really lend themselves to making larger volumes of data available to collaborators. Where such a requirement exists, there are two typical paths being followed currently – as follows:

Internal sharing within the same department is primarily undertaken via creating shared areas on a network so that all parties have the ability to access.

Sharing of data with external collaborators or those in other University departments follows a number of different routes, including:

- Via secure websites set up specifically for such tasks
- SSH File Transfer Protocol / Secure File Transfer Protocol
- Copying data onto physical disks and shipping between parties via courier
- The use of file sharing software such as Bit Torrent Sync and Wetransfer
- Exchanging of USB drives / sticks at meetings and conferences

As the amount of research data being generated continues to grow and the focus on collaborations increases, the need to share large data sets with external parties is likely to escalate. Whether this can be facilitated via a centrally-provisioned 'Dropbox-type' solution or will need to be supplemented by other options *(such as shipping data on physical disks)* is a challenge that needs to be considered. Based on feedback received, the sheer size of the data sets sometimes shared is likely to require UIS to offer a disk in / disk out service at the data centre, where data can be copied onto a HDD and shipped via courier to another party (*and also the reverse – where data is received on disk from another party and needs to be removed*).

5.6. Sharing of data with students

As well as using solutions such as Dropbox for file synchronisation and sharing with their research collaborators, some PIs mentioned that they were increasingly being used as a means of communicating with their students. This was seen as a more effective means of sharing documents than via email – especially when email quotas were often a subject of growing concern. Whilst the amount of storage space that would be required on a University-supplied 'Dropbox-type' solution by students should be minimal *(compared to a PI),* we should nevertheless ensure that the student population are also considered in our plans.

5.7. Use by administrators

Rather than focus solely on the Research community, a limited number of administration users were also included in the interview process. Interestingly, Dropbox-type solutions were also being used by administration staff to manage specific tasks – such as the dissemination of documentation to PIs for review and decision-making as part of the Student Admissions process.

If / when the University decides to introduce its own Dropbox-type service, it should seriously consider making it available to administration staff as well so that it can support their interaction with other users. In fact, rather than be selective and only offer such a solution to specific role types, there is benefit in making it available to everyone - across all roles and all Schools, Faculties, Departments and Colleges.

5.8. User interest in a UIS Dropbox-type service

Based on the feedback received, there is massive support for the potential introduction of a Dropbox-style service by UIS. Aside from a couple of researchers who were fully invested in using Dropbox Pro and didn't believe there was any real benefit to them in adopting another solution, pretty much everyone else was very interested in the prospect *(with many trying to push for further details of potential features, quotas and a launch date)*. The only other exceptions were a couple of Post-Doctoral Research Associates who were also currently using Dropbox Pro as a means of storing and backing up all of their data, and particularly liked the fact that they were not tied to the University in any way - and would still have full access to all of their data if/when they moved on from Cambridge at some point in the future.

Although the potential for inclusion in a proof of concept / pilot was not a question raised during the discussions until fairly late in the project, everyone who was asked about this indicated they would be very happy to participate - which underlines the considerable appetite for such a solution.

On the basis that interest in UIS's future plans is very high right now as a result of this initiative, we need to make sure that we maintain ongoing communications with those who participated in the project and keep them abreast of future intentions. So as not to lose momentum, the publication of some sort of high level 'roadmap' at the earliest possible opportunity would be of great benefit - so that people can use it to make informed decisions when planning their next steps from a storage perspective.

5.9. User requirements for a UIS Dropbox-type service

As so many of those interviewed are already using Dropbox and love its rich features and ease of use, the bar has already been set very high for us. We are told that offering users anything less than they have currently will be regarded as a failure and will severely limit the levels of adoption. There is definitely an appetite for a UIS-provided alternative to Dropbox however, but for it to be widely adopted as a solution, it needs to:

- Provide all of the features that people are now so used to *(e.g. sharing/ synchronisation / versioning)*

- Have sufficient storage capacity for 'free' that for the majority of users there is no need to pay for an upgraded service. Should also offer a 'paid for' option for users that require a larger quota.
- Have a very high level of reliability and availability, with clear support processes and SLAs.
- Support both internal *(to Cambridge)* and external sharing.
- Be highly intuitive and very easy to use, and provide a pleasant user experience.
- Include some on-line help for those who are less comfortable with technology and might require some instruction and hand-holding to begin with.
- Not be restricted to particular user types. For it to be truly effective it should be offered to all users – including Admin staff and Students.
- Be hosted in the UK *(ideally at a University of Cambridge-owned data centre – not in the cloud.)*
- Easily accessed – ideally via single-sign on / Raven authentication.
- Many users require tablet / smartphone compatibility, so should have a mobile app as well as a web & desktop client version.
- Have additional *(optional)* security features – e.g. ability to set an expiry date on a link as well as add a password/pin code.
- Ability to share at both a file and folder level. Able to easily see who has access and amend as needed.
- Allow synchronisation across multiple devices and platforms. Support selective sync. so that users are able to decide exactly which folders are synchronised.
- Fully auditable - for example able to see who has uploaded / downloaded a file *(and when)*
- Maintain a full version history of files. Allow access to previous versions as well as deleted files.
- Support in-app document editing so that users don't need to download, update and upload again *(This is something Dropbox doesn't offer – which is why some are using GoogleDrive instead).*

In summary, what users are demanding is a fully-featured Dropbox-type solution that is very easy to use; is hosted at Cambridge; has more capacity than existing 'free versions' and doesn't cost anything. *(To clarify – the feedback provided is that we should aim to provide a reasonable amount of free space to everyone – with the option to pay for more capacity if they need it).*

5.10. User types and suggested quotas

There are two typical user types as far as Dropbox-type solutions are concerned – the 'heavy user' and the 'light user'. Heavy users account for around a third of researchers, with the remaining two-thirds being light users currently.

- **The 'heavy user'** A third or so of researchers interviewed subscribed to a 'paid' service in order to access additional storage capacity, with the vast majority of these opting for Dropbox Pro. This came at a cost *(circa £8 per month)* – which they were happy to bear in the absence of a viable alternative. A couple of PIs reported that as well as having personal accounts, they also funded such solutions for their whole research groups. Even though these 'paid' options typically provide access

to 1TB of space, many stated that they weren't using anywhere near their full quota right now. Of these heavy users, a very high proportion (~85%) would consider switching to a University-provided solution so long as there was a compelling reason for them to do so. *(Either cheaper for the same features and space, or the same cost for more features/space).*

- **The 'light user'** The remaining two-thirds were making use of 'free' services – sometimes subscribing to more than one solution in order to maximise the amount of free cloud storage space available to them. Although the basic Dropbox service only provides 2GB of free storage, many had more than this *(10GB or more was not uncommon)* through introductions and participation in the 'Dropbox Space Race'

When questioned as to how much the University should ideally provide in terms of 'free space' on a Dropbox-type solution, the answers varied wildly and ranged between 5GB and 50GB. With 30,000 or more potential users to consider however, such a variance clearly has a huge impact on overall storage capacity, so offering a range of options might be a sensible approach.

Where a larger quota is needed, users have indicated they are happy to pay for this privilege *(and already do so in many cases on solutions like Dropbox Pro),* so an option to allow an increase to any 'basic' free provision should be made available. The current gap between a typical free service and the paid for alternative is vast *(in the case of Dropbox, it leaps from 2GB to 1TB)*, so a series of interim steps would be welcomed – particularly as some of the subscribers to 'Dropbox Pro' stated that they don't use the full 1TB available to them.

One option that we may want to consider is to provide all users with access to a modest free quota to begin with *(maybe 5GB?),* but give them the ability to 'top up' – as and when their demands grow. Allowing users to obtain access to a larger quota *(maybe in multiples of 50GB or 100GB?)* for a small additional fee should satisfy everyone by providing them with the flexibility to tailor their quotas and only pay for the space they require.

Providing such a high degree of flexibility might prove to be a real administrative headache however, so an alternative approach to consider is to offer a series of 'fixed' service options – as depicted below:

Service	Quota	Cost
Bronze	5GB	free
Silver	50GB	£1 per month
Gold	250GB	£3 per month
Platinum	1TB	£7 per month
Diamond	5TB	£25 per month

NOTE: these quotas and costs are suggested for demonstration purposes only and might not be viable.

Regardless of the final approach selected, it is very clear that 'one size fits all' won't work as far as space on a Dropbox-type solution is concerned, so we need to consider how additional capacity can be made available where required – whilst ensuring minimum administrative overhead for all parties concerned.

6. Archiving and long term data retention

Archiving is a complex subject and one where there appears to be a fair bit of confusion right now. Some researchers don't fully appreciate the difference between backup and archiving – and in some cases used the two words interchangeably. There are also conflicting priorities and viewpoints involved – dependent on whether you are looking at things from an IT perspective; as a Researcher or as a Sponsor.

6.1. Data Retention - Sponsor Requirements

Researchers were specifically asked how long their project sponsors required them to retain data for post completion of their research, and what needed to be retained. The vast majority *(>80%)* of respondents were rather unsure as the exact duration of retention and stated that they would need to check back on the terms of their grant to make absolutely sure. This would suggest that data management and retention is not something that is at the forefront of researchers' minds, and is something that is typically dealt with at the end of the grant *(if at all)*.

None of those interviewed had ever been required to prove to their sponsor that data was being retained in line with their requirements, nor had anyone been formally audited in this regard. As a result, many researchers are fairly relaxed right now in terms of delivering what is being demanded by their sponsors from a data retention perspective.

Most believed the minimum retention period to be somewhere between 5 and 10 years, and independent investigation into a handful of the most common sponsors supports this – as summarised in the table below:

Sponsor Name	Minimum Retention Period
AHRC	Min 3 years after end of grant.
BBSRC	10 years
CRUK	5 years
ESRC	"openly available to a maximum extent possible"
EPSRC	10 years
MRC	10 years
NERC	10 years minimum. 20 years for major importance outputs
NIH	3 years
STFC	10 years
Wellcome Trust	Min 10 years

Sources: http://www.sherpa.ac.uk/juliet/ http://researchdata.ox.ac.uk/funder-requirements/

Whilst there is a lot of useful advice and guidance on the University's Research Operations Office web pages in terms of managing research projects, there doesn't appear to be anything on there related to data retention requirements and the need to draft a Data Management and Sharing Plan as part of the grant submission, so the information summarised above has had to be obtained from alternative sources. It is suggested that such content be considered for future inclusion on the Research Operations Office pages in order to make it clearer as to exactly what is required. Rather than having to re-invent the wheel, a really good example of such information is provided by the University of Bath – where very comprehensive details of funder requirements for data management and retention are

available on its website – as per link here. Similarly, University of Oxford has also published some useful information which might be of interest and serve as a good starting point.

The majority of publicly-funded research requires that as well as publishing the research outcomes, that PIs also make available the raw data and other supporting materials *(e.g. the code used to generate the results)* where it is appropriate to do so. This is intended to stimulate new investigations and analysis outside the scope of the original project and to maximise the value of the research initially funded by making it available to the wider scientific community.

A couple of sponsors *(NERC and ESRC)* currently provide their own repositories for long term publishing and retention of data, whereas others make it the responsibility of the research institution to provide an appropriate solution.

Since 2002, the University Library has offered DSpace as an institutional repository for the publication of research outputs and theses. It can additionally take some supporting data, but this is reviewed on a case by case basis and charges for deposit may apply - at a one-off fee of £2000 per TB *(excluding VAT).* It is not known how widely this service is being used however – owing to the fact that it is very difficult to extract statistics from the system.

Further investigation (via re3data.org – the Registry of Research Data Repositories) shows that there are a huge number of research data repositories available currently. At the time of writing, there are over 1,000 separate repositories worldwide, of which 150+ are in the UK. An average of ten new repositories is currently being added each week, which would suggest that the publication of research data is highly fragmented right now – and is not showing any signs of slowing down. A few of those interviewed suggested that there might be some opportunities for future rationalisation in this space, but were not aware of any current plans to do so.

6.2. Data Retention – Researcher Requirements

When questioned as to how long they personally wanted to keep hold of data for *(regardless of any sponsor or regulatory requirements)*, the resounding response from researchers was 'forever'. Many still carried forward data on their PCs and laptops from 15 or 20 years ago – even if they are very seldom referred back to again in most cases. Others had data on a whole variety of media *(including various tapes, disks, CDs, DVDs and USB drives)* which were often stored in boxes on office shelves or locked away in cupboards and bottom drawers - 'just in case' they might be required again in future. It is widely accepted by researchers however that the odds of being able to read some of their older media are fairly low, so whatever solution UIS might offer in terms of longer term storage needs to consider data deterioration and physical media obsolescence – especially now that some sponsors are requiring data to be retained for up to 20 years *(and researchers want to keep things forever…)*

Only a very small minority reported that they revisited aged data on a fairly regular basis *(typically to continue to build on the initial research).* For the most part, older data was hardly ever or never re-accessed and was only kept as a precaution – just in case it might be useful again.

Such practice was likened by one researcher to putting things he no longer used up in his attic. Whilst he was comforted by the fact that he hadn't disposed of something which might be required again at

some point, his attic was overflowing with items which hadn't been touched in many years. Even if he did need something again in future *(which was deemed to be unlikely),* he would really struggle to find it in amongst all of the other things that had been consigned to the loft – and also had no guarantee that it would still be in working order.

When IT staff were asked how far back data goes on their departmental storage, many still had files which were created up to 20 years ago – the majority of which had not been accessed in a very, very long time. A small number of departments *(circa 20%)* had an unofficial policy of asking users to move / delete older records when space was running low, whereas in most cases it was deemed far easier to simply keep on adding more storage as space ran out - as the older data only took up a relatively small proportion of the overall capacity and wasn't really worth worrying about too much. Although this model can work for the medium to long term in those areas where data volumes are modest and growth rate is fairly steady *(e.g. Arts & Humanities),* it isn't sustainable for departments or research groups where volumes are already considerable and the rate of growth is high.

6.3. The current approach to archiving

Regardless of the requirements of sponsors, those interviewed believed that a proper mechanism for archiving of data for the longer term was a key requirement, but none had any sort of formal archival policy or solution in place right now.

The current approach to 'archiving' falls neatly into two groups – as follows:

The **'DIY'** approach. After the end of a research project, data was copied onto DVDs, tapes or local hard drives for the Principal Investigator (PI) to manage themselves on an ongoing basis. Data was then deleted off the departmental storage in order to free up valuable space. This approach was either initiated proactively by the PI as part of their project completion activities, or more often than not was instigated reactively by the IT department at a later date when storage capacity was running out and they leaned on PIs to free up some much-needed space.

The **'Do Nothing'** approach, where it was considered far easier to simply add further storage capacity as and when required and to carry on storing everything as 'live' data indefinitely.

Most of those interviewed currently fell into the 'do nothing' camp, but owing to the way in which data volumes are ever-increasing, there was very strong support for the need to do something different so far as longer term storage was concerned.

From those departments visited, around 10 or so had expressed some interest in the Arkivum offerings, and two *(Cambridge Enterprise and Dept of Chemistry)* had gone as far as attending Arkivum demonstrations and were seriously considering moving discussions on to the next level.

6.4. Researchers concerns about archiving

Whilst it was agreed that doing nothing in terms of longer term data retention and archiving is not really a viable option, there were a few concerns which were regularly raised by researchers – as follows:

- How much is it going to cost them? As well as having concerns about being able to afford to archive initially, there were also concerns about being able to afford to pay for the next 10, 15, 20 years - or however long the data should be retained for. Research grant income for PIs is often quite 'spiky', so the longer term affordability needs to be considered.
- How much time and effort will they need to spend on describing their data *(i.e. adding metadata)* as part of the archiving process – and how much of this can be automated?
- If sufficient effort is not spent on capturing metadata, how can they be sure that they will be able to easily identify their data again in future?
- Will they be the ones who will decide on what data *(and when*) is committed to archive, or will this be an automated process outside of their control that runs in the background based on pre-defined rules?
- What happens to the data if/when they leave Cambridge? Will they still be able to access? If not, who will become the new 'data owner'?

NOTE: Although initially anticipated as being a potential concern for PIs, fast speed of re-access to archived data was not a key requirement – with most stating they would be happy to wait minutes or hours *(and in a couple of extreme cases even up to a few days)* for data to be available to them again.

6.5. Metadata

Many of the researchers interviewed have stated that the definition of good metadata is key to future data retrieval, but have also said that they don't want to spend too much time and effort on creating this metadata, so ideally a solution needs to be found where as much as possible can be captured automatically, with most of the remainder being optional so that those who are prepared to invest additional time and effort on their metadata can still do so.

This view is echoed by the Digital Curation Centre (DCC) which states: *"while data curators, and increasingly researchers, know that good metadata is key for research data access and re-use, figuring out precisely what metadata to capture and how to capture it is a complex task"*

It would appear that there are already dozens of metadata standards in existence – many of which are discipline-specific, so the complexity of assigning metadata should not be underestimated and needs to be a serious consideration when defining a longer-term archiving solution.

6.6. Future approach to archiving

Based on the feedback provided, there appear to be two separate solutions needed when it comes to 'archiving' requirements.

Solution A - A solution is needed for data that the researcher wishes to keep for the long term for 'personal use', but where there is no need to share it *(for example with external parties as part of an 'Open Data' requirement).* This type of requirement would be best served by providing an option of cheaper / slower storage in a central location – as opposed to on faster/more expensive local storage. This approach should require minimal effort on behalf of the researcher and no assignment of any metadata. Interviewees have already stated that although they want to keep data 'forever', much of

it is seldom *(if ever)* referred back to again so speed of re-access is not that important. Although some initially demanded re-access speeds 'as fast as possible', further prompting unearthed the reality that delays of minutes *(or even hours/days)* were quite acceptable. On the basis that much of their older data was already 'cold' and sitting on shelves or in cupboards on tapes; CDs/DVDs; external hard drives and de-commissioned desktops and laptops, a delay in being able to re-access was agreed to be perfectly reasonable after all.

Solution B – There is also need for a 'proper' archival solution which will enforce the capture of metadata and support the potential to make such data available to others – both internally and external to the University. This requirement would be best served by a 'DSpace'- type solution and needs to be capable of supporting future requirements for 'Open Data'. Requested features for such a solution as follows:

Mandatory

- The number one requirement was that it must be affordable. It is acknowledged that such a solution is unlikely to be made available for free, so if there is a charge involved, a "p*ay once, store forever"* model would be preferred so that total lifetime costs are known up-front and can be factored into budgets *(and passed onto sponsors).*
- Be easy to access. Ideally linked to Raven to negate the requirement to remember additional usernames and passwords.
- Be capable of long term storage *(i.e. for 20 years or more),* and provide support against data degradation and media obsolescence.
- Have clear ownership and access permissions so that data security is tightly managed. Allow for additional data manager(s) to be added, so that data can be accessed after the original owner has left *(e.g. retired or left the University)*
- Have formal support processes and a service level agreement.
- Have at least two copies of the data in separate physical locations. (3 ideally!)
- Be ISO27001 compliant
- Require a minimal amount of mandatory metadata to be added, but allow users to optionally add further metadata where required.
- Should support persistent URLs (or DOIs) so that the data may be referred to and linked from published papers.

Desirable

- Some projects require data to be deleted after a certain time, so it would be highly desirable to be able to enter a 'keep until' date *(or leave blank for indefinite storage).* Once this date has been reached, data owner should be emailed to ask if they want to extend or wish for data to be deleted.
- Allow data owner to amend the 'keep until' date *(or remove it totally so that data kept indefinitely)*
- Issue recurrent email reminders *(suggest annual)* to data owners so that they can check that their data still needs to be retained and take action where needed.

Note: It is understood that there is a need for the University to deliver something to support '**Solution B**' as soon as possible, as EPSRC are demanding that for all papers published after 1st May 2015, *"that data relied on in published research findings will, by default, be available for scrutiny by others. The expectation could be satisfied by citing such data in the published research and including in such citations direct links to the data or to supporting documentation that describes the data in detail, how it may be accessed and any constraints that may apply. It is important that any such links are persistent URLs such as DOIs."*

7. Backup – current state

The vast majority of departments visited provided a backup capability for any data stored on their central filestore. Notable exceptions were: Physiology, Development & Neuroscience; Material Science & Metallurgy and Chemical Engineering & Biotechnology because they had no central departmental storage provision available.

IT interviewees reported that they often reminded researchers that their local data (*e.g. data held on laptop and desktop HDDs)* were not backed up centrally and that users needed to make separate arrangements to protect against data loss. Discussions with researchers identified that such arrangements currently included:

Do Nothing One of the PIs interviewed had no backup capability whatsoever in place. All of his data resided on his desktop hard drive and was not copied anywhere else.

Time Machine & Time Capsule ~15% of researchers relied on Time Machine and a Time Capsule to back up their Macs. This was a solution that was very well liked by those who used it.

Cloud backup Approx. 15% relied on cloud-based solutions as a means of backing up their local data. Most commonly used was used Dropbox Pro, but other solutions such as Carbonite were also mentioned.

Syncing between devices Half a dozen or so interviewees opted to sync. data between multiple devices as a means of backing up. Rsync was the most common method of doing so, with Goodsync also being used by one individual.

Email Quite a few of those interviewed griped about the fact that they were often reaching their email quotas *(particularly those more established researchers who had been at Cambridge for many years),* yet sending emails to themselves as a means of 'backing up' key files on an ad-hoc basis was a practice that some researchers favoured.

USB drives/sticks By far the most common method of backing up local data was via the use of external local storage – typically USB drives, USB sticks and in one case even SD cards. Where such storage was being utilised, the process of backing up was very often manual and required the researcher to initiate it *(as and when they remembered to do so).* The 'backup' copy was typically kept in the same location as the primary copy however – with just a couple of those interviewed stating that they took their backup copy home in the evening for safekeeping.

Saving to central dept. storage Where departments provided researchers with space on central storage, some used this on an ad-hoc basis to copy selected data from their desktop/laptop for safekeeping. This practice sometimes caused frustration for the IT departments however – particularly where users created a new 'backup folder' on each separate occasion – resulting in multiple instances of the same data being stored and tying up valuable space.

Despite the rather fragmented approach to backup of local data, very few researchers reported having experienced any significant data loss as a result of loss/theft/failure. On the contrary, a couple mentioned instances where they had had their laptops lost or stolen and had been able to recover their data again within hours through the use of Time Machine.

In a similar vein, none of those in an IT-related role had needed to invoke a major recovery from backup in a full-blown 'disaster recovery' situation, and more often than not any recovery task needed was merely to restore a file or folder which had been deleted by a user in error.

8. Backup – future requirements

Based on the feedback received, there are two areas where UIS should offer services from a backup perspective – as follows:

Department backup At a departmental-level, there is interest from IT users in being able to utilise space in a data centre for off-site backup purposes. This is an option that many might consider *(subject to the usual caveats of cost and performance)* when it came to replace or upgrade their current backup solutions.

Individual backup At a research group *(or even individual researcher)* level, there is considerable interest in the potential introduction of a 'managed backup service'. As well as providing off-site storage capacity, such solution would also need to include the relevant software and provide a simple interface where users can manage their own backup schedules and retrieve data themselves as and when required.

9. Regulatory Compliance and Data Security

Interviewees were asked whether or not they needed to comply with any specific regulations or standards with regards to data management – and in relation to ISO27001 in particular. Very few had even heard of ISO27001, let alone had any awareness as to whether or not they were required to comply with it. The assumption was that if they didn't know anything about ISO2700, that it didn't apply to them. Just 3 individuals believed that it was a current requirement.

Those involved in medical research made it very clear however that there were a number of restrictions in terms of the way that sensitive data *(e.g. patient identifiable data)* should be handled. The Clinical School Computing Service (CSCS) is already very well versed in all of the requirements and regulations involved in handling sensitive data, and offers a Secure Data Hosting Service (SDHS) in order to adhere to these regulations.

It is understood from discussions that the use of cloud-based storage solutions are prohibited for patient-identifiable data owing to the potential risk of 'offshoring' sensitive data. Others cited a need to comply with Principle 8 of the Data Protection Act – which states that personal data shall not be transferred to a country or territory outside the European Economic Area (EEA). Solutions hosted in a UK datacentre *(such as the West Cambridge Data Centre)* would be perfectly acceptable to them however.

A couple of those interviewed sometimes handled animal-related data – and again had some concerns about the security of cloud-based storage solutions but were unable to highlight any specific policy or standard that detailed exactly how such data should be managed. A few researchers handled data that was commercially sensitive and required a high degree of security. In a couple of such instances, the external collaborators insisted that the University used tools and methods of secure data exchange which they themselves provided – rather than allowing any other means to be used.

For the vast majority of interviewees however, data security was something they didn't really worry too much about. There was a general uneasiness about the potential lack of security of cloud-based storage solutions *(such as Dropbox and GoogleDrive)* as data was stored outside of the UK, but they were still being extensively used nevertheless. Nearly everyone interviewed would be far more comfortable with a University-hosted solution instead – particularly if such solutions allowed them the option to add further security precautions where appropriate – for example:

- Ability to set a password / pin number that users need to enter prior to accessing a link.
- Ability to set an end date on a link, so that it automatically expired.

In summary, whilst very high levels of security were not a requirement for most interviewees, it is clear that an option for provision of physically separate storage *(potentially in a secure cage or locked room with restricted access)* might be required by a small minority, so this is a service that we should include as a standard offering.

10. General Concerns

Whilst all interviewees were very supportive of UIS's plans to offer central storage solutions, as one might have expected there were a handful of concerns raised too – which may or may not be valid but have been summarised below nevertheless.

- **Cost.** This is the number one concern – and will have the biggest impact on whether or not UIS services are widely adopted. Aligned to this are worries about potential future loss of access/service. As research grant income can be highly irregular, what happens if income dries up and a PI can no longer afford to pay? Does their access get removed?

- **Network performance.** This was raised on a few occasions, and there was a concern that unless UIS increased network speed alongside the introduction of new storage solutions, they might not deliver the required level of performance.

- **Single datacentre.** Some of those on the Addenbrookes campus believed that the West Cambridge Data Centre was too far away, and suggested that a secondary data centre closer to them would make any new services proposed by UIS far more attractive. A few individuals raised concerns over contention – particularly if very high numbers of individuals were attempting to access the same storage location. Having two or three datacentres in locations chosen to optimise access speeds for users was suggested as a means of addressing this.

- **User experience.** This concern was raised primarily in the context of a Dropbox-type service, but could be extended to cover any new services delivered by UIS. A good user experience was listed as a high priority by many interviewees, so we need to spend time and effort on usability *(as well as functionality).*

- **Proper support processes & SLAs.** Users have become highly dependent on the various file sharing / synchronisation solutions that they use, and expect *(and currently receive*) a very high level of service. As well as being able to deliver levels of availability to match existing offerings, we need to have formal processes and SLAs in place to support this.

11. Manners of access to storage

Interviewees were specifically asked which manner(s) of access they would require of central storage. Those in an IT-related role were far better positioned to respond to this and requested the following methods:

- NFS
- SMB / CIFS
- SSH / SFTP
- Block Level
- Web
- API

Most of the interviewees in a 'non-IT' role did not really have any particular view as to how storage should be accessed – all they wanted was to be able to 'see' it from their desktop/laptop as an extension to their existing drive structure and have access to a simple 'drag and drop' user experience.

12. Availability

As far as availability is concerned, all researchers stated that they required very high levels of availability - ideally 24 x 365, as their work didn't follow the typical *'9 to 5 with an hour for lunch'* model. Justifying this, many stated that as the 'normal' day was often spent teaching they ended up performing research activity during evenings and weekends. Burning the midnight oil to achieve funder deadlines, as well as collaborations with people in different countries and time zones also meant that access to data at all times of day and night was often needed.

It is accepted however that there could be occasions when solutions might not always be available – owing to the need for maintenance/upgrade. So long as any such unavailability was planned and well communicated a few days in advance this was not perceived to be a show stopper for the vast majority of interviewees. It was requested that such downtime should never exceed the planned 'window' however and should be kept as short as possible. Performing such maintenance outside of 'core UK hours' was also a requirement – with the early hours of Monday morning being the time that was regarded by many as being least obstructive.

The only area which specifically demanded 100% availability was the University Library – owing to the fact that much of what they do is public facing and attracts world-wide attention. For the Library, any unavailability of public-facing services would be totally unacceptable.

13. Authentication and authorisation

Ease of use and accessibility are very high on the agenda for all researchers – particularly when it comes to a Dropbox-type solution. Whilst such solutions clearly need to be accessed both inside and outside of Cambridge, there is a strong desire for them to be as easy as possible to access for University users– with Raven authentication being stated as the preferred option for internal users so that there is no need for additional sign-on.

Where files and/or folders are shared via a Dropbox-type solution with external collaborators, the ability to require an additional passcode/pin number to be able to access a link was requested as an *(optional)* additional security measure. Whether or not to require such additional authentication should be decided by the 'owner' of the file/folder at the point of sharing.

Exactly how services might be accessed by external parties is a subject that requires a great deal of further consideration. For example - in order to comply with research sponsors 'Open Data' requirements, an archive solution might need to provide restricted access to named individuals *(e.g. whilst under embargo)* – as well as support unlimited public open access. How such access should be managed would need to be part of the detailed design of such a solution and is far wider than being just a technical challenge.

14. Metrics

Extracting specific reporting requirements from those interviewed was a difficult task – especially whilst being very careful not to end up 'leading' the discussion in any way by making suggestions in order to try to stimulate ideas. This was the one area where very limited input was received – regardless of whether the interviewee sat in the IT or researcher camp. Where any requirements were stated, these predominantly related to a Dropbox-type solution and addressed the following questions:

- What is my total quota and how much is used *(or still available?)*
- How has my usage changed over time? *(trend of last 3 / 6 / 9 / 12 months)*
- File / folder history *(when created? who by? date last updated?)*

- File / folder permissions *(who is it shared with currently? who was it shared with previously?)*
- When was a file viewed / downloaded & who by?

One of the main reasons why IT interviewees struggled with the definition of specific reporting requirements was that their future role in terms of supporting / administering such solutions was not clear. Owing to a lack of clarity as to whether or not local IT staff had any part to play, coming up with potential reporting requirements was very challenging for them.

The way in which future services are managed and charged for will be a key influencer in defining reporting requirements, so rather than focusing too much on trying to second-guess reporting needs at this early stage *(when roles and responsibilities are unclear)* , it might be best to consider reporting alongside the definition of policies and procedures for ongoing management instead.

15. Performance

Fast storage is not a requirement in the vast majority of cases. Most *(>85%)* of those researchers interviewed are interested only in outright storage capacity rather than requiring any sort of high performance. The minority of researchers who did have higher performance storage needs *(e.g. those involved in DNA sequencing, imaging or bioinformatics)* were either already using the High Performance Computing Service *(HPCS),* and/or had a local capability which satisfied their requirements. Such users were very keen for this to remain the case and would not consider moving their high performance storage needs elsewhere, as it was important for them to have such storage within very close proximity to their other equipment. Once performance was no longer a requirement however, they were happy to consider moving data to another location for longer term retention.

Those who currently used the High Performance Computing Service *(HPCS)* were very complimentary about the knowledge and helpfulness of staff and the services they provided – with many stating that they intended to make greater use of such services in future rather than continue to expand their local capabilities.

Network speed was mentioned as a major consideration by half a dozen or so participants – mainly from those departments who are a reasonable distance from West Cambridge *(e.g. Addenbrookes).* Many IT interviewees from the Addenbrookes campus have also suggested that services delivered from another datacentre local to themselves would be a more attractive proposition than services delivered from the West Cambridge Data Centre. Whether or not such concerns about network performance are valid is unknown and would require further investigation.

Contention issues were also raised as a potential concern by a couple of parties – particularly as uptake of services increases. Again, it is not known whether or not such concerns are valid.

16. Related projects

Through the course of the interview process, some other projects have been identified which are very interested in the outcomes of this activity. Dr Clare Dyer-Smith *(Co-ordinator for the Cambridge Big*

Data Strategic Initiative) and Dr Marta Teperek *(Co-ordinator for Open Data)* would welcome a copy of the final report and have requested that they also be kept abreast of any ongoing developments.

Similarly, there is a project currently underway between the Clinical School and the Biological School with a view to forging a 'Joint School Computing Service'. Paul Sumption *(Biological School)* and Ling-Yan Lau *(Clinical School)* are both involved in this initiative and have asked to be kept abreast of the outcomes.

17. Other services

During the course of this project, a small number of those interviewed raised unsolicited suggestions for additional services which they would be interested in receiving - if UIS were to offer them. Whilst this input may not be directly related to this project, it is recorded here so that the feedback is not lost.

Service Desk application Owing to the highly devolved nature of IT delivery across the University, there is no common Service Desk application in use. In fact, many of the smaller departments do not use any system at all for managing their workload, and instead rely on email inboxes to track activity. A couple of the Computer Officers in such departments suggested that if UIS were to provide them a Service Desk application (such as 'Request Tracker'); they would be very happy to use it. NOTE: The reason that Request Tracker (RT) is specifically mentioned is that this appears to be quite widely used across the University already.

Source Code Repository A small number of the researchers interviewed (~5) mentioned that they write a lot of code to support their research activity and rely heavily on source code repositories such as Bitbucket and GitHub in order to manage this code. A centrally-provided 'University GitHub' would be welcomed by such individuals.

Hosted Rack Space There was interest from a couple of parties in the potential of renting rack space *(with power & connectivity)* from UIS. This is an option that the Clinical School Computing Service already offers to Addenbrookes users. Assuming that this is a service that UIS is also happy to offer, including it as an option in a service catalogue/price list might be a good idea as it might encourage others to put some of their existing equipment into the West Cambridge Data Centre as their first step towards centralisation.

18. Conclusions

Based on the feedback gathered from ~120 interviews, it is very clear that there is considerable interest in the potential introduction of solutions based on a shared storage infrastructure. As previously mentioned, these include things like:

- A Dropbox-type solution.
- Backups – both as a 'managed backup service' for individuals *(or groups)* and also via provision of 'off-site' storage space for departmental backups.

- A cost effective means to support long-term retention of data that is seldom *(if ever)* re-accessed and does not need to be shared with others.
- A 'proper' archiving solution – i.e. one that supports the 'Open Data' requirements of research sponsors by making data publicly available.
- Provision of general storage space – which departments *(or even groups / individuals)* can buy into as and when required – either to supplement and/or replace current local storage provision. NOTE: Options for scratch space as well as backed up storage would need to be provided.

Whilst departments *(and groups / individuals)* have a high degree of flexibility available to them and are free to decide how they satisfy their ongoing storage demands, the option of buying into clearly-defined services from the University Information Service is regarded as being highly attractive – so long as they are cost effective.

In terms of priority, the provision of a Dropbox-type solution generated by far the greatest interest and should ideally be tackled ASAP – not only because it touches the greatest number of users *(30,000 + potentially),* but also owing to the fact that we might be exposed to data security risks in this space currently.

A managed backup service is also seen by many as a being a high priority owing to the risk of data loss that some groups are exposed to right now.

External pressure from sponsors to comply with the requirements of 'Open Data' means that the introduction of an archiving solution to support this is also a matter of high importance.

Regardless of which order such services are addressed in, it is key that UIS shares its plans and intentions *(via some sort of roadmap possibly?)* as soon as possible after the end of the initial requirements capture activity. This will allow people to make informed decisions as to how they proceed with satisfying their future storage demands - and might encourage many to delay local expenditure and opt for central provision instead.

Appendix A – List of participants

The following individuals generously gave up their time to partake in the study and share their views and requirements, and their input is greatly appreciated.

Aside from a couple of exceptions, all input was received via face-to-face meetings – which took place between 3rd November 2014 and 10th February 2015. Some supplementary questions were also answered by email or telephone.

Name	Department / Faculty / Institution
Katherine Abell	Department of Chemistry
Richard Adams	Department of Physiology, Development and Neuroscience
Boris Adryan	Department of Genetics
Jun Aizawa	Department of Earth Sciences
Jethro Akroyd	Department of Chemical Engineering & Biotechnology
Paul Alexander	Department of Physics (Astrophysics)
Raad Al-Rawi	University Information Services
Mark Ansell	University Information Services
Robin Ansell	Department of Chemical Engineering & Biotechnology
Rob Anthony	MRC Cognition and Brain Sciences Unit
Gavin Alexander	Faculty of English
Paul Alexander	PI in Physics
Jenny Barna	Department of Biochemistry
Daniel Bates	Faculty of Law
Andrew Baughan	Corpus Christi College
Peter Benie	Department of Engineering
Dave Berry	Wolfson Brain Imaging Centre
Mike Bithell	Department of Geography
Richard Black	Department of Radiology
Alan Blackwell	Computer Laboratory
Richard Bramwell	Institute of Criminology
Barney Brown	Office of External Affairs and Communications
John Burnham	University Information Services
Jonathan Chin	Isaac Newton Institute for Mathematical Sciences
Ian Clark	Department of Genetics
Graham Christie	Department of Chemical Engineering & Biotechnology
Paula Coelho	Department of Genetics
Ian Cross	Faculty of Music
Gabor Csanyi	Department of Engineering
John Davey	Department of Zoology
Paul Dampier	Registrary's Office
Tim Dickens	Department of Chemistry
Adam Dickinson	MRC Epidemiology Unit
Lila Dimanti	Haematopoietic Stem Cell Lab (CIMR)
Alastair Downie	Gurdon Institute
Clare Dyer-Smith	Big Data Strategic Research Initiative
Steven Eglen	Department of Applied Maths and Theoretical Physics (DAMTP)
James Elliott	Department of Materials Science & Metallurgy
Vin Everett	Cambridge Institute for Medical Research (CIMR)
Ian Frame	Department of Earth Sciences

Bruce Fraser	Faculty of Classics
Dr Laurent Gatto	Department of Biochemistry
Andrew Gerrard	Faculty of Law
Tony Gledhill	Department of Materials Science & Metallurgy
Matthew Gratian	Cambridge Institute for Medical Research (CIMR)
Lesley Gray	University Library
Gillian Griffiths	Cambridge Institute for Medical Research (CIMR)
Allegre Hadida	Judge Business School
Rebecca Hannah	Haematopoietic Stem Cell Lab (CIMR)
Richard Harrison	Department of Earth Sciences
Robert Hay	Department of Physics
John Hill	Department of Physics
Leona Hope	Department of Physics
Jo Howson	Department of Public Health and Primary Care
Peter Humphreys	Stem Cell Institute
Roger James	Department of Haematology
Steve Jones	MRC Cancer Unit
Steve Kimberley	School of Arts & Humanities
Julian King	University Information Services
Markus Kuhn	Computer Laboratory
Brian Lam	Institute of Metabolic Science (IMS)
Ling-Yan Lau	Clinical School Computing Service
Christopher Lester	Department of Physics
Bob Longman	University Information Services
Ian Luff	Department of Psychology
David Lyness	Department of Earth Sciences
Peter MacCallum	Cancer Research UK Cambridge Institute
Ezequiel Martín	Department of Medical Genetics
Epaminondas Mastorakos	Department of Engineering
Sean McGuire	University Information Services
Stuart Meacham	Department of Haematology
Gos Micklem	Department of Genetics / DAMTP
Paul Miller	Judge Business School
Eric Miska	Gurdon Institute
Iain Morrison	Department of Chemical Engineering & Biotechnology
Stephen Mounsey	Department of Engineering
Leila Muresan	Cambridge Advanced Imaging Centre (CAIC)
Mark Neal	University Information Services
Sheilagh Ogilvie	Faculty of Economics
Kevin O'Holleran	Cambridge Advanced Imaging Centre (CAIC)
Ole Paulson	Department of Physiology, Development and Neuroscience / BCNI
Willem Ouwehand	Department of Haematology
Andy Parker	Department of Physics
Craig Peacock	Faculty of Economics
Nicholas Perentos	Department of Physiology, Development and Neuroscience
Jennifer Pollard	School of Arts & Humanities
Andrew Raine	MRC Mitochondrial Biology Unit
Jason Randall	Clare College
Carolyn Read	School of Clinical Medicine

Sylvia Richardson	MRC Biostatics Unit
Chris Rogers	Statistical Laboratory
Mike Rose	Department of Applied Maths and Theoretical Physics (DAMTP)
Megan Sammons	Computer Laboratory
Aylwyn Scally	Department of Genetics
Oliver Shorttle	Department of Earth Sciences
Debora Sijacki	Institute of Astronomy
Brian Simpson	University Information Services
Paul Sisodia	Department of Radiology
Matt Sims	MRC Epidemiology Unit
Matthew Skipper	Institute of Criminology
Wen-Tao Song	Department of Physiology, Development and Neuroscience
Emmanuel Stamatakis	Division of Anaesthesia
Tony Stevenson	University Information Services
Howard Stone	Department of Materials Science & Metallurgy
Tim Stuart-Buttle	Centre for Research in the Arts, Social Sciences & Humanities (CRASSH)
John Suckling	Department of Psychiatry
Paul Sumption	Stem Cell Institute
Alex Sutherland	Institute of Criminology
Elizabeth Swann	Centre for Research in the Arts, Social Sciences & Humanities (CRASSH)
Marta Teperek	Registrary's Office
Andrew Thomason	Department of Pure Maths and Mathematical Statistics (DPMMS)
Adam Thorn	Department of Chemistry
Lacey Wallace	Faculty of Classics
Nick Wareham	MRC Epidemiology Unit
Lorenz Wernisch	MRC Biostatics Unit
Guy Williams	Wolfson Brain Imaging Centre
Mark Williamson	Department of Chemistry
Jonathan Wilson	Cambridge Institute for Medical Research (CIMR)
Mark Wilson	Department of Paediatrics
Hayley Woffendin	Department of Oncology
Nigel Wood	Department of Physiology, Development and Neuroscience
Michael Yardley	Cambridge Enterprise