Data management questionnaire results : IDMB Project

1.1 First Iteration (8th – 31st May 2010)

In order to obtain quantitative information about research data management across a spectrum of areas, including current practice, policy, and governance, an online questionnaire was devised based on the approach used in the previous DataShare project. This adapted and implemented the Data Audit Framework.[i1] The project team refined the questionnaire to provide a balance of level of detail, completion time, and categorisation of results in the most usable format. The questionnaire comprised 30 questions and took an average of 9.6 minutes to complete.

The questionnaire was split into three sections: About You; About your data; Final comments[2]. While most of the questions were quantitative, additional text boxes for qualitative information were included where appropriate.

The survey and interviews required ethics clearance before they could be launched, and they were initially completed by participants in the project pilot disciplines; Humanities (Archaeology only), Electronics & Computer Science (ECS) and Engineering Sciences.

Due to the ethics clearance, individual invitations were sent out to potential respondents to the questionnaire website (http://www.isurvey.soton.ac.uk/663) and the URL was advertised on the project website.

Note that the answers to the questionnaire were from individuals, and the statements they make based on their own knowledge. Therefore some answers may not reflect the true situation, only the respondent's view.

1.2 Second iteration (May 2011)

Senior stakeholders agreed that the questionnaire should be rolled out across the whole University - however in practice returns were uneven across faculties. See Table 1 for a breakdown of the combined responses to both iterations by discipline and research role. Any summary statistics relating to the whole dataset have been weighted by the response in each discipline.

One of the larger responses was received from Electronics and Computer Science – this increased the response rate from 47 to 81 when both sets are combined (excluding the small number who responded to both iterations). The results are similar to before, with no significant differences.

The largest response was from Humanities – this significantly increases the response (from 34 to 76) when combining both iterations of the survey. In addition while the original survey exclusively covered Archaeology, almost all of the new responses were from other research areas in the faculty. The latter cover the other research areas in similar depth (English, film, history, music and philosophy) with a better response from modern languages. However the

[1] Southampton Data Survey: our experiences and lessons learned http://www.disc-uk.org/deliverables.html

[2] For details of the questionnaire and dataset this report is based on see http://eprints.soton.ac.uk/195959

proportion of staff and researchers responding to the survey was much lower than for Archaeology.

A reasonable response was received from Social Sciences (mostly from social statistics/demography and sociology and social policy, with some response from the other areas). In addition there were results from other disciplines in the Faculty of Social and Human Sciences: Psychology, Mathematics, Geography and Education. Although Social Sciences is much larger (and thus a higher response is expected), the small numbers of these results make them harder to draw conclusions from (especially Education).

	Academic or equivalent (e.g. HEFCE- funded)	Research Fellow (e.g. Project-funded)	Research Student	Other	Total
Electronics and Computer Science	35	12	33	1	81
Engineering Sciences	10	11	11	2	34
Geography	3	2	4	2	11
Humanities	46	8	16	6	76
Maths	5	1	2	0	8
Psychology	2	1	2	2	7
Social sciences	12	4	6	0	22
Total	113	39	74	13	239

Table 1 - Survey respondents, broken down by school and role (Q1.2 and Q1.3)

1.3 About You

This section describes the responses for each question, including additional qualitative data where appropriate

Q1.3 Research Role

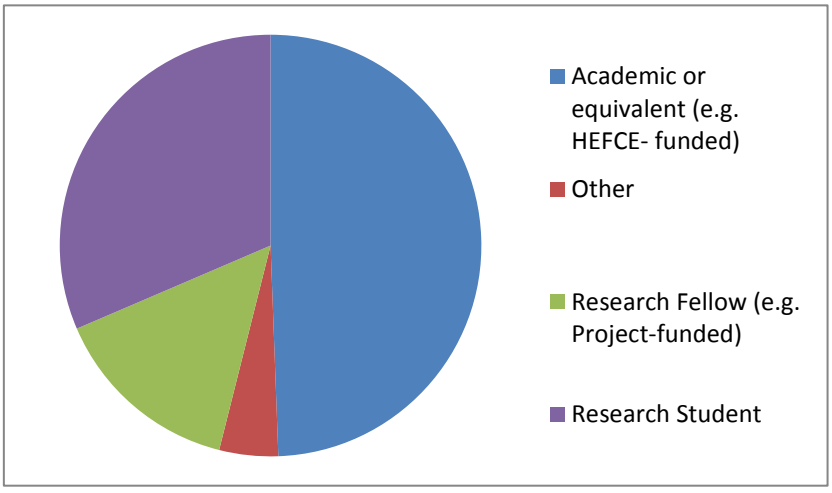

Figure 1 - Please describe your research role

A handful of respondents identified other research roles:

- Publicity coordinator (ECS)

- Research Engineer on KTP [Knowledge Transfer Partnership] Project (Engineering Sciences)
- Administration; Experimental Officer; Pedagogic research; Subject Centre (Humanities)

Q1.4 Area of Research

In most cases the range of responses reflected the breadth of research activity across all research groups[3]. For a few cases responses were disproportionately received from a small number of areas:

For Humanities 44% of the responses were from Archaeology (the only area surveyed in the first iteration), 21% from Modern Languages and a fairly even response from the remaining areas.

For Social Sciences 54.5% of responses were from Social Statistics and Demography, 22.7% from Sociology and Social Policy and an equal response from other areas.

1.4 About your data

Q2.1 Ownership of research data

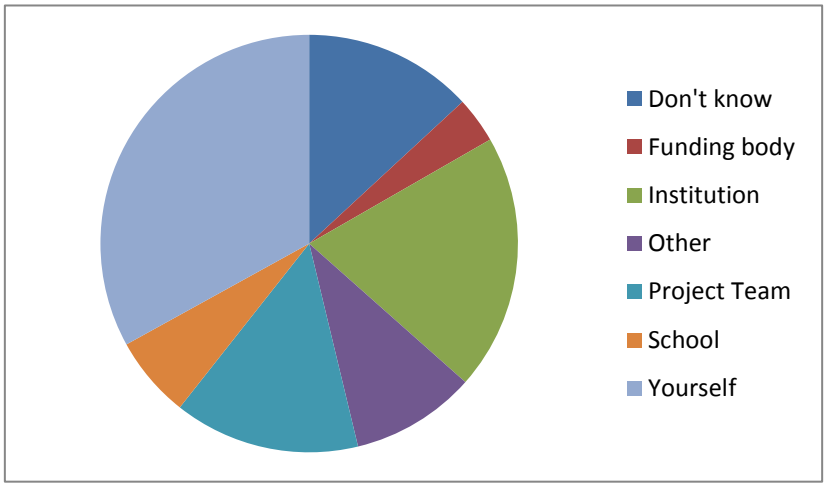

Figure 2 - Who do you believe owns your research data?

The responses for this question, show that 33% identify themselves as owning their data. This compares with 26.2% for School/Institution (combined) and only 3.6% for the funding body. 13.1% of respondents did not know who owned their research data.

A number of specific funding bodies were identified as owning data including PASCAL Network, Airbus France, Rolls Royce, Defence Science and Technology Laboratory, EPSRC and the BBC.

There were a number of comments indicating the ownership of their data varied depending on project agreements and could be shared between multiple bodies. A comment about archaeological objects indicated that they are the property of the owner until an agreement is reached for deposit in a museum

[3] See the quantitative data summary for more details at http://eprints.soton.ac.uk/195959

Q2.3 Data types

The responses to this question show different patterns depending on the discipline. There are similar quantities of observational, experimental, computer code and derived data in engineering, ECS and Maths. This is perhaps expected as there is typically a data pipeline going from experimental/observed/simulation data to derived data.

Humanities has the highest proportion selecting 'other', this is much higher (28.2%) if Archaeology is excluded. They identified less well with the characteristics given and more with types of data (photographs, pieces of music, etc.).

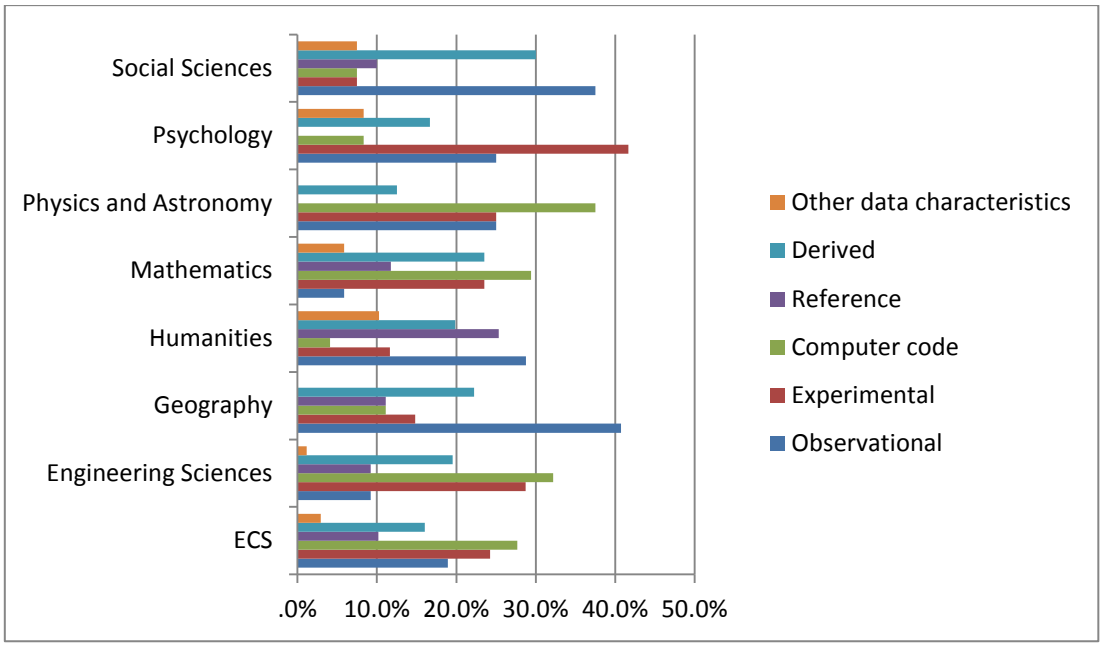

Figure 3 - Characteristics of the data – please tick all that apply

The spread of different data types is wide, indicating that it is difficult to focus on any one type. Interestingly, 27% is office productivity file formats (e.g. Microsoft Word, Excel and PowerPoint, or equivalent). Also, media formats account for 21.3% - e.g. images, audio, video. There are differences for each discipline – e.g. Engineering and ECS have significantly more computer code than other areas; Social Science and Psychology have significantly more SPSS files.

Respondents from Archaeology identified a number of additional data types - Physical objects (lab samples, artefacts), site drawings, ArcGIS files, bibliographic data, Illustrations based on data, current ethnographic data and 3D motion capture. ECS identified medical trial data, Mathematica [computational/modelling software]; Organisational data; calculations; rules & scoring systems; interviews and case studies.

(Some of these were identified under responses to 2.3 Characteristics of the data.)

Q2.4 Hardware/Software compatibility issues

Only 16.5% of respondents to this question highlighted compatibility problems with their data.

Of these, data on floppy disks and zip disks were commonly identified. Other issues included Betacam Video tapes of interviews, paper-based data (ECS), audio cassette tape,

1930s-1950s glass lantern slides, 1950s large plate negatives, files in older versions of AutoCAD/Geoplot (Humanities).

Q2.5 Experiences of reusing data from previous projects

66.4% of respondents said they did re-use their own data from previous projects. A mixture of comments indicating this was easy, difficult or varied; with a majority (58-82%) in each school indicating reusing their own data was relatively easy. The only exception was Psychology at 14.3% - based of a fairly small sample (7).

In Humanities there were three comments about difficulties in using old data with modern software, and also concerns about loss of data and accessibility to others. Also one respondent mentioned visiting museums to refresh his memory of objects (with varying success in locating them).

In Engineering Sciences difficulties in using electronic data over 10 years old and time needed to process data were identified. In ECS need for format conversion and the importance of reusing code was identified (including the use of versioning software e.g. "Forge").

Q2.6 Have you ever used data from external sources?

59% of respondents said that they had used data from external sources, with Psychology (28.6%) and Engineering Sciences (45.2%) being significantly lower, and Geography (81.8%) and Social Sciences (85.7%) being higher.

Q2.6a Experiences of using external data

This qualitative follow-up to question 2.5 elicited interesting comments. Respondents' experiences of using external data varied widely, from being seamless, to being extremely difficult and potentially expensive.

ECS, Humanities and Social Sciences had significant usage of free data, and issues of cost and licensing inhibiting use of other data. More specific issues included licensing for software associated with specific instruments and for reproduction rights.

Humanities also identified issues with format of the data, lack of standards in 3D data, time consuming process (and organisations slow to respond to requests for data) and good documentary/manuscript data available from Record Offices.

ECS mentioned use of open source software, difficulties in getting hold of/using data and lack of support from data providers.

In Engineering Sciences less had used open access data (availability was a problem), more had to contact the author, and sometimes to ask them questions to be able to use it.

Q2.7 Who is responsible for managing your data

There was general consensus that the individual researcher was responsible for data management – 63.9%. Two national data centres were named (PASCAL NoE, Archaeology Data Service). ECS and Engineering Sciences also identified research students and sometime researchers from another school as being responsible. One researcher in Modern Languages indicated the Library was responsible.

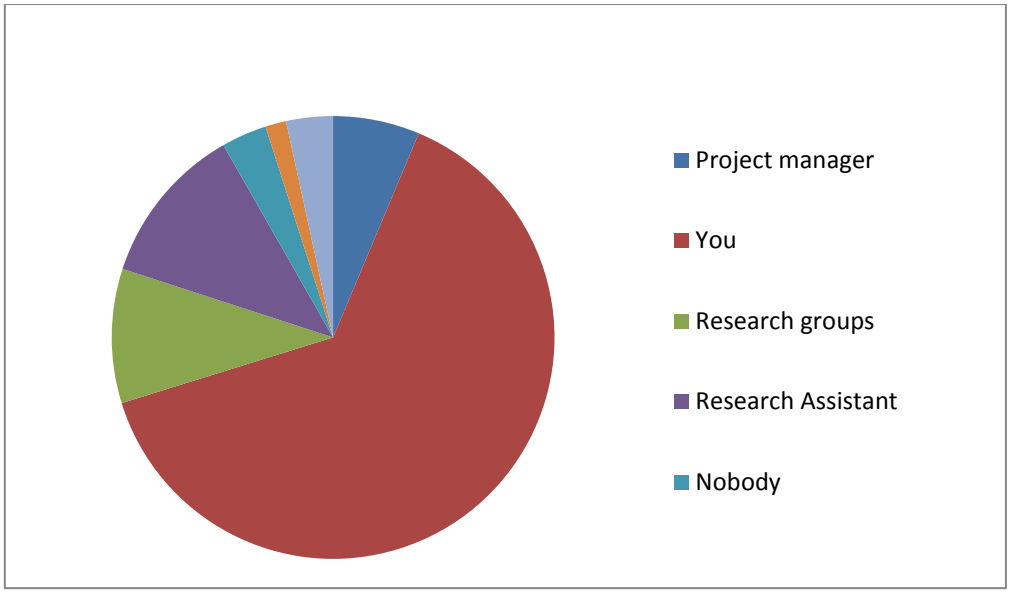

Figure 4 - Who is responsible for managing your data? - please tick all that apply

Q2.8 Where do you store your current data

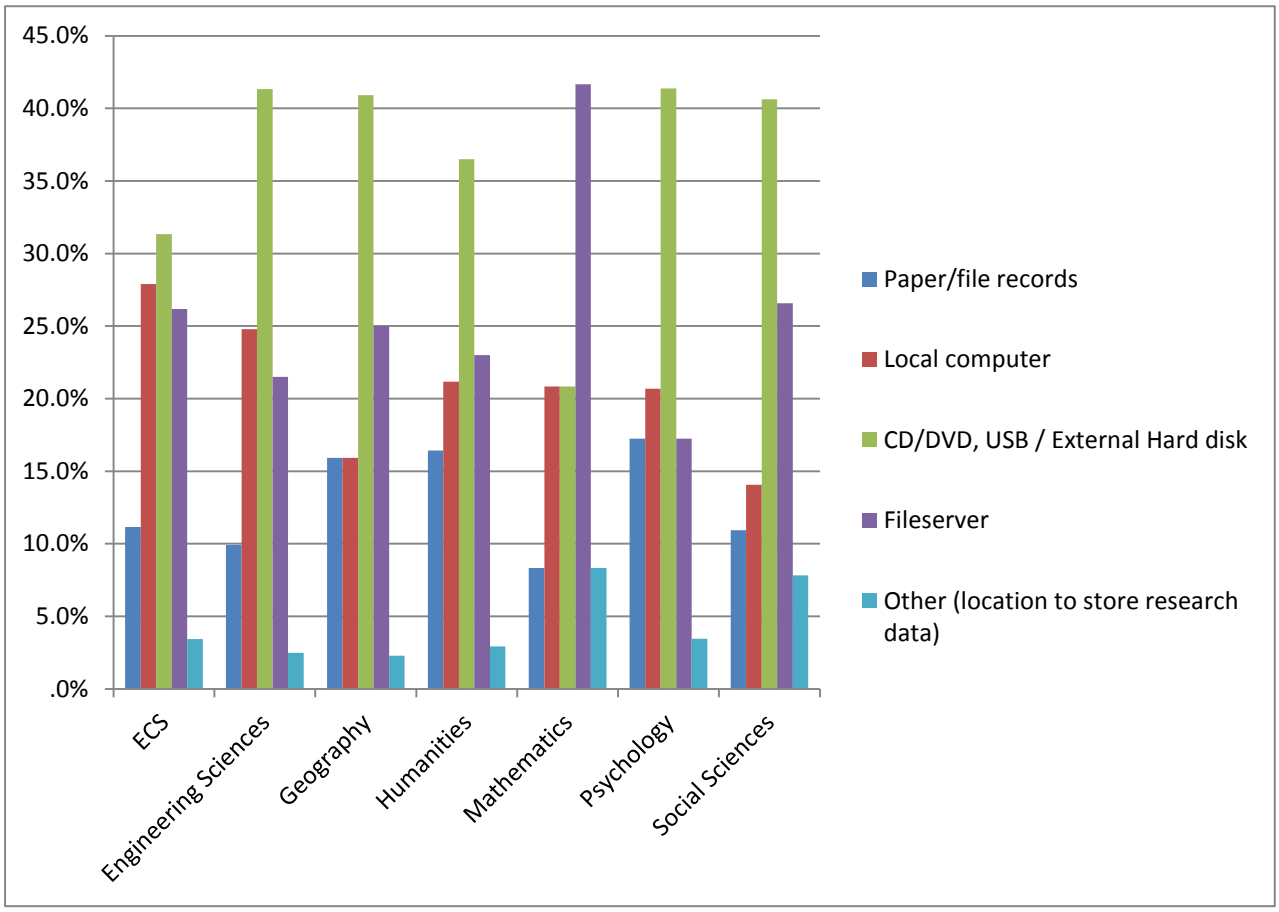

Figure 5 - Where do you store your current data? - please tick all that apply

A wide array of data storage locations were highlighted in this question. 24% use their local computer, with 34.9% using CD/DVD, USB flash drive or an external hard disk. 24.3% of respondents used a file server, either at the University or off-site. There is some variation by discipline but the overall there is a fairly similar picture.

People in ECS use a number of external/commercial storage facilities including source forge, Google code, Google Documents, Zotero filestore and a website provided by a sponsored company. Also mentioned included a remote collaborative environment server hosted by the Open University and private clouds.

There were many respondents from ECS who used file servers provided by their school. In addition they mentioned: Forge SVN server (using the git interface), a compute server which runs both relational and RDF-based data stores, IT Innovation file servers, PASCAL Forge/EPrints/Data Repository/Video Lectures, servers bought for the project and those run by students in their group.

Engineering Sciences use external facilities including HECToR, National Grid Service, network drives of industrial partners and some crucial codes on Amazon's S3 service. School provided facilities included SES Research Folder, "Guide" shared drive, Spitfire Cluster (run by iSolutions), RIfI, MS home server 2TB, SESNET *fs1* file server and CT data held locally on servers bought on research grants. Some used their home PCs for backup.

Archaeology mentioned using Flickr, Integrated Archaeological Database (YAT hosted). A number of school-provided shared areas were mentioned, but appear to all refer to Resource (centrally provided by iSolutions). Note a recent audit of archaeology filestores indicates that the situation is more complex.

Q2.9 How much electronic data do you currently retain?

This question was split to try and ascertain data requirements for typical projects, and over the career of a researcher. We present detailed and aggregated results in Table 2 and Table 3 respectively.

It is interesting to note that for a typical project the breakdown by School indicates that in Humanities 53.1% of respondents held up to 100GB, with only 6.3% holding more than 100GB. Also, significant numbers of respondents in Humanities, Maths, Psychology and Social Sciences did not know how much data they held.

	ECS	Engineering Sciences	Geography	Humanities	Maths	Psychology	Social Sciences
Up to 100 GB*	74.2%	45.2%	71.4%	89.5%	100.0%	66.7%	100.0%
Over 100 GB*	25.8%	54.8%	28.6%	10.5%	0.0%	33.3%	0.0%
Don't know	4.3%	3.1%	0.0%	35.9%	25.0%	50.0%	35.3%
N/A Paper /film based data	1.4%	0.0%	0.0%	4.7%	12.5%	0.0%	0.0%

Table 2 - Data stored per project, including aggregated results

* Percentages exclude respondents who didn't know how much data they held (or with paper/film based data).

The percentage with over 100 GB is a good indication of the storage needs of the discipline. Again there was a significant proportion of don't knows in Humanities, Maths, Psychology and Social Sciences.

	ECS	Engineering Sciences	Geography	Humanities	Mathematics	Psychology	Social Sciences
Up to 100 GB*	61.4%	28.6%	60.0%	63.2%	100.0%	50.0%	78.6%
Over 100 GB*	38.6%	71.4%	40.0%	36.8%	0.0%	50.0%	21.4%
Don't know	10.8%	3.4%	9.1%	32.8%	25.0%	42.9%	25.0%
N/A Paper/film based data	1.5%	0.0%	0.0%	1.7%	0.0%	0.0%	5.0%

Table 3 - Data stored in total per researcher, including aggregated results

* Percentages exclude respondents who didn't know how much data they held (or with paper/film based data).

Q2.10 How long do you keep your data?

The working practices of researchers, in terms of their curation and preservation behaviour, were questioned here. The most significant result is that 45.9% of respondents state that they keep their data forever. This question does not include correlated data on where this data is kept, but read in line with Q2.8, it can be assumed that a significant number of users do this locally (local PC, CD/DVD, external hard drives, USB flash drives).

It shows that researchers value their research data, and therefore prefer to keep it well beyond the end of a typical project (assuming most projects last less than 10 years).

	ECS	Engineering Sciences	Geography	Humanities	Maths	Psychology	Social Sciences	Total
Up to 10 years	31.9%	37.5%	27.3%	11.6%	12.5%	14.3%	20.0%	23.5%
Over 10 years	15.3%	9.4%	18.2%	14.5%	0.0%	57.1%	10.0%	14.4%
Forever	34.7%	31.3%	36.4%	60.9%	62.5%	14.3%	55.0%	45.9%
Don't know	18.1%	21.9%	18.2%	13.0%	25.0%	14.3%	15.0%	16.2%

Table 4 - Longevity of data storage

Q2.11 Do you keep data for compliance reasons?

Only 8.7% of respondents stated that they had to maintain data for compliance reasons. There was more retention of data for compliance in Psychology (57.1%), Social Sciences (20.0 %) and Geography (27.3%), but from fairly small samples.

A supplementary question was asked to ascertain how long data was held for compliance reasons. The most common length for keeping data to maintain compliance was 5 years; also other periods were given included 15 years, until the project is over, and forever. Also recycling data for other projects and keeping data to 'cover our back' was mentioned.

Q2.12 Have you ever experienced storage problems due to the size of the files?

A significant proportion (40-50%) of users said that they had experienced storage constraints at some point, Engineering Sciences being particularly affected (71.9%), understandable since they had the most data. Maths was much lower (25%), and Psychology a little higher (57.1%) – which also correlates with the amount of data held (see above). Social Sciences is an exception – less data is held yet they experience the same amount of storage problems as ECS.

Q2.12a Method to overcome storage issues

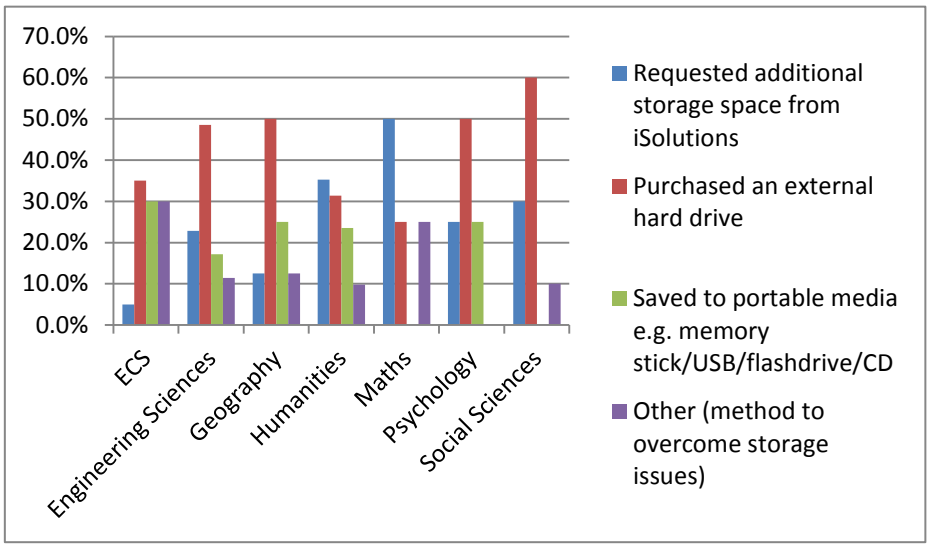

Figure 6. How did you overcome these storage issues? (please tick all that apply)

In order to overcome storage constraints 60.4% sought a local solution (external hard drive, CD, USB flash drive) to overcome this. However there is some variation by discipline - requesting additional central storage from iSolutions was the preferred route in Maths, and was significant in Humanities and Engineering Sciences.

ECS respondents often identified requesting additional space on ECS servers and improving capacity of existing servers. Also using external storage and removing oldest data were mentioned.

For Engineering Sciences there were comments identifying deleting/compressing files, an extra internal hard drive and storing data at home.

In Archaeology, school specific storage space and storing on older departmental computers were mentioned.

Q2.13 How frequently do you backup your data?

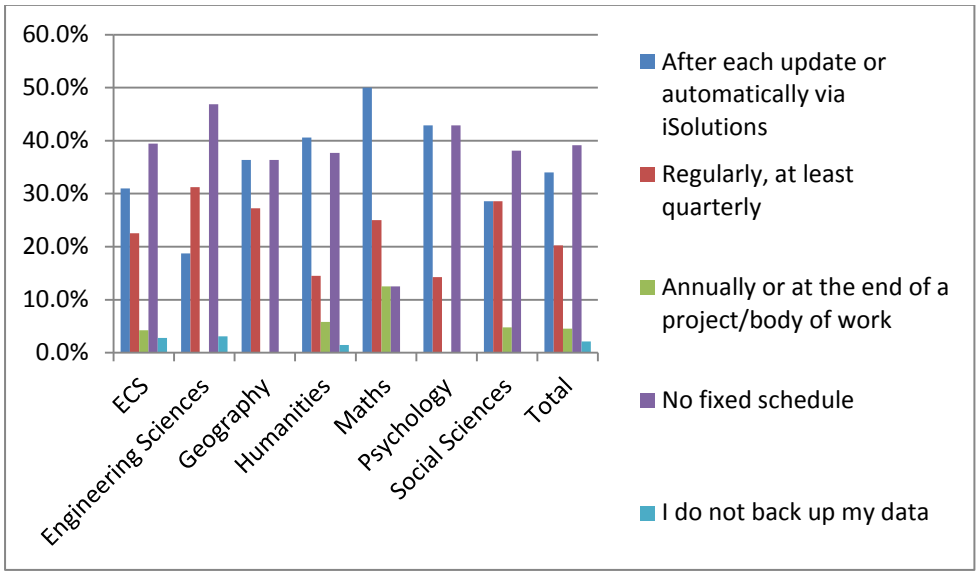

Figure 7. How frequently do you backup your data?

Backup behaviour shows that over half (54.2%) of respondents had a regular backup routine (at least quarterly), with 14.3% used the University central backup system via iSolutions.

Only a very small fraction (2.1%) did not say they performed backups. Over a third (39.2%) said that they did perform backups, but not according to a regular schedule.

Q2.14 Where do you back up your data

This question permitted multiple responses. Over two thirds (64.4%) used a local solution to store a backup of their research data. There is some variation by discipline, and increased usage of print-based backup in Social Sciences, Humanities and Psychology.

ECS mentioned using Google, public/private clouds, duplicate file servers, source code backed via ECS subversion repository, IT Innovation internal file servers (NAS and SAN based systems).

Engineering Sciences mentioned using local Microsoft Home Server, company backup system, emailing to themselves (web email account), home PC and Amazon S3 cloud storage. Archaeology mentioned using Integrated Archaeological Database, Windows Live Skydrive and external partners' server systems.

	ECS	Engineering Sciences	Geography	Humanities	Maths	Psychology	Social Sciences	Total
Local System	65.3%	73.9%	75.0%	62.4%	54.5%	68.4%	55.3%	64.4%
Server System	28.1%	18.8%	25.0%	27.4%	36.4%	26.3%	28.9%	27.2%
Other	4.8%	5.8%	0.0%	1.1%	9.1%	0.0%	7.9%	3.2%
Paper/file records	1.8%	1.4%	0.0%	9.1%	0.0%	5.3%	7.9%	5.2%

Table 5. Where do you back up your data? - please tick all that apply

Q2.15 Depositing data with other services, such as the UK Data Archive

There was significant variation by discipline. Note in Humanities there is a higher response from Archaeology (28.1%) than the rest of the faculty (9.5%). This is probably due to the existence of the Archaeology Data Service. Although there is a similar service for History, there are not comparable services for many areas of the humanities since the closure of the Arts & Humanities Data Service.

ECS	Engineering Sciences	Geography	Humanities	Maths	Psychology	Social Sciences	Total
2.8%	0%	9.1%	18.3%	0%	28.6%	28.6%	10.1%

Table 6 - Do you deposit your data with other services, such as the UK Data Archive?

The main services identified were the Archaeology Data Service (ADS) and UK Data Archive (identified by Social Sciences, Humanities, Geography and Psychology). In addition the Africa Centre for Health and Population Studies and the Max Planck Institute for Demographic Research were identified by Social Sciences, and TPTP (Thousands of Problems for Theorem Provers) Library and github.com by ECS.

Q2.16 Keeping track of where data is stored, and what it relates to

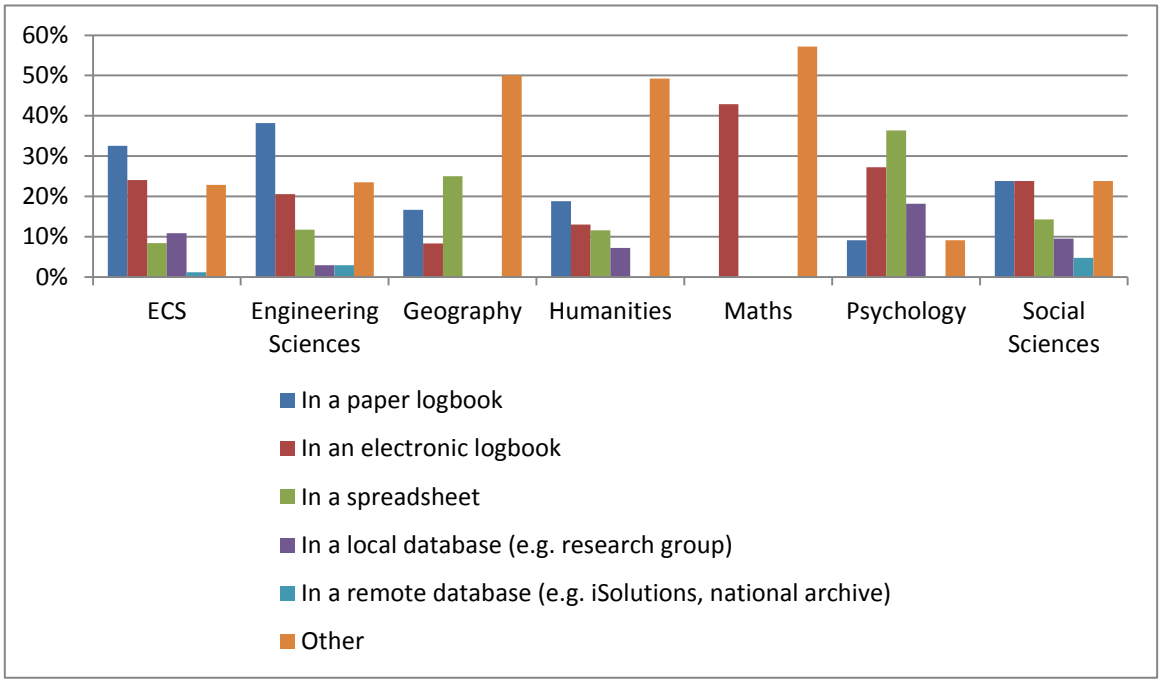

Figure 8. How do you keep track of where your data is stored, and what it relates to? - please tick all that apply

Almost a third (27.1%) of respondents used a paper logbook for keeping track of their research data. Just over another third used either a spreadsheet (10.4%) or electronic logbook (19.6%) to perform this task. Only 9.7% used either a local or remote database for this.

33% of researchers responded with *Other*. The most common other solutions identified were a mental record and through file/folder naming, with the following also mentioned:

- Giving each project a unique code, paper/computer files cross reference these

- Coloured pdf
- Electronic written reports
- Keeping track of location of paper data sheets by emailing to themselves
- Using specific programs / services (SPSS, National Grid Service Oracle Service, arcgis metadata file, PASCAL NoE)

There were also many comments saying their files were not well organised.

Q2.17 Allowing others access to your data - during the lifetime of a project

ECS	Engineering Sciences	Geography	Humanities	Maths	Psychology	Social Sciences	Total
75.7%	78.1%	72.7%	63.6%	62.5%	57.1%	63.2%	70.5%

Table 7 - During the lifetime of a project, do you allow others access to data on which you are working?

Three-quarters of respondents did say that they allowed access to their data to other people during a project. The most common answer was colleagues (sometimes including students), collaborators and industrial partners. A few share data to those that request it. A few in ECS host public datasets, and one in Archaeology uses the GENIE data management archive. One in Engineering Sciences uses their own data as teaching aids.

For those who answered *No*, reasons for this were solicited. The most common (42.4%) was that sharing was not required. Confidentiality or data protection issues were significant (29.5%), with some license agreements expressly prohibiting sharing of data (9.4%). A small proportion (11.9%) of respondents also stated that that the data was not fully documented.

Other issues with sharing this data identified were - data is available as required by local research team; not wishing to give access to data they had paid for and no-one had asked for it.

Q2.18 Allowing others access to your data - after a project has ended

ECS	Engineering Sciences	Geography	Humanities	Maths	Psychology	Social Sciences	Total
76.5%	74.2%	63.6%	73.5%	50.0%	42.9%	50.0%	73.8%

Table 8 - Do you allow others to access your data once the project is finished?

Again, just over three quarters of respondents (73.8%) shared their data after the end of a project. Confidentiality and data protection was again highlighted as a significant reason for not sharing (44.8%), along with license agreement prohibiting sharing (15.8%) and lack of documentation (14.5%).

Geography, Maths, Psychology and Social Sciences appear less willing to share data after a project has finished than during it – and are generally less willing to share than Engineering. They may be a different understanding of the question and more reliance on depositing in external repositories. This could also reflect the research culture in the disciplines concerned.

Similar to Q2.17 colleagues, collaborators and partners were common answers. More willingness to share data with external users, mostly on request in Engineering and ECS, with mention of depositing in ESRC archives in Social Sciences.

Other issues identified were mostly sharing not being required or no-one had asked. As in 2.17 one respondent was unwilling to share data they had paid for. One person in Archaeology indicated data was available via publications, and objects in a local museum.

Q2.19 Details of data management plans requested by funder(s)

Only 9.7% of respondents said that they had been required to submit a data management plan to their funders. This was 5.3% for Social Sciences. Although the requirement for ESRC funding has been introduced relatively recently, many staff may not yet have had experience of these. The response for Engineering Sciences was 6.3% where the EPSRC do not have such a requirement.

Humanities commonly identified AHRC as requiring this, and also mentioned the British Academy, ESRC and MEPF/ALSF (Marine Environment Protection Fund). Engineering Sciences mentioned RIfI (Research Institute for Industry - industrial consultancy unit in the school). ECS mentioned NERC, JISC and the DTC (Doctoral Training Centre). Geography mentioned NERC and MRC.

Q2.20 Details of School policies

It is clear from this question that most researchers are unaware of the existence of School data preservation policies (91.3%). Engineering Sciences mentioned the RIfI file storage policy, and the importance of storing data on a research group server to preserve IP. In ECS a data disposal policy, and informal policies in IT Innovation (e.g. two copies, two technologies, and two locations) were mentioned, in addition to the Data protection Act. Geography mentioned ethical considerations for interviews, and one respondent in Social Sciences referred to University-level guidance.

Q2.21 Would you find it useful to have university wide guidelines to manage and maintain your research data?

ECS	Engineering Sciences	Geography	Humanities	Maths	Psychology	Social Sciences	Total
47.2%	68.8%	40.0%	60.0%	12.5%	71.4%	52.4%	54.3%

Table 9 - Would you find it useful to have university wide guidelines to manage and maintain your research data?

The majority of respondents (54.3%) agreed that university-wide guidelines for managing research data would be useful. However there is significant variation by discipline.

Q2.22 The biggest problem with regard to managing and storing your data

This open-ended question was aimed at finding out the most important problems facing researchers. Commonly identified issues were:

- Problems in backup due to
  - lack of space
  - time consuming (i.e. not automated)
- Organising data
  - finding and keeping track of their data (including knowing how an image/spectra/etc. was obtained),
  - issues with version control especially for code.
- Lack of space on fileservers (mostly an issue for Engineering Sciences/Archaeology),
  - problems caused by large files (storage, processing). Hadoop cluster + hdfs storage was identified as a solution for the latter by one person.

ECS also identified the need for better guidance for good practice and issues with remote access, not all data being publishable, developing bespoke systems to manage data, limits on data crawling in public data sources (delegate a crawl across numerous IP addresses to get round this) and not using iSolutions for support.

Archaeology also identified a need to make data accessible after the project is finished. Other issues in Humanities included loss of knowledge when key staff leave, lack of support for research websites and difficulties using the Institutional Repository.

Geography also identified ethical issues of storing confidential data on laptops during fieldwork that need to be kept secure.

Social Sciences and Psychology identified lack of storage space for print data.

Q2.23 How can the University make data management and storage easier for you

Question 2.23 aimed to solicit suggestions on how the institution can assist researchers with their data management. The main issues identified were:

- More guidance needed (but not rigidly imposed rules – diverse needs of researchers) – including what facilities/services were available, data management training
- A need for more automated backup, especially for data on desktop/laptops
- More storage space on network drives (20/100 GB/unlimited) and better ability to access remotely. Comments about lack of space for PhD students, enterprise workers and having a high capacity my documents/known archive space for each person
- Some in ECS, Engineering Sciences and Archaeology requested an '*EPrints for data*'

ECS respondents also identified a need for a *git* versioning server (with enough hard disk space), software to help with data management/storage and a service to scan lab note books.

Archaeology identified needs for uploading data to the University while in the field, developing an ethnographic/archaeological archive, more space for physical data ability to archive data to the ADS. Humanities also identified the need for better technical support and facilities for scanning.

Engineering Sciences also identified problems with accessing My Documents and using memory sticks on library computers, and need for a system to zip files and archive logically (Linux), providing external hard drives and initial advice on data management.

1.5 Further Comments

Q3.1 Any further comments

The questionnaire concluded with an open question to capture respondents other thoughts. These are broken down by School below.

Humanties

- Happy with current set-up - large amount of backed-up space (J drive)
- Would be helpful to have a reference to query on data ownership/advice on sharing primary data
- I keep records of meetings etc rather than actual research data
- Space to store paper data is an issue (and the money or time to photocopy to keep backups). We have taken to taking photographs of workbooks in the field so that we have some electronic back-up of the day's work.
- The main issue for me is about deposition and long-term curation of (digital) photo and video archive

- What to do with paper archives
- All my interview data are confidential. ESRC data from a previous project have been made available to their Database in confidential anonymous paper form as agreed.
- I think that data management needs a higher priority -- it's not something that seems to be particularly well publicised/catered for in Humanities.
- No really problems with this in my view. Confidentiality and ethics makes it impossible share data

Electronic & Computer Science

- Large server needed for storing video data
- Run training session "how to manage your data resource"
- Issues in preserving AV data with complex rights issues limiting retention/use - completed DRAMBORA risk assessment for project (threats including storage systems, file formal/bit-level preservation e.g. migration) - including total cost of ownership over 10 years
- Main issue for old data where author has left the university, need to avoid introducing more admin
- Avoid a complex centralised bureaucracy - just lots of storage with good backup, everything else is optional
- Questions in survey pitched more to empirical data researchers thus harder to answer for modellers

Engineering Sciences

- Suggestion to provide all PhD students with 1-2 TB external hard drive
- Need for 4 TB of backed up space

Geography

- Data management is something everyone is expected to do well but no-one is ever told how to do it well so often data can end up in a mess. I have always thought basic training on data management including file naming conventions storage file hierarchies backups and basic data management and manipulation within applications (i.e. excel) would be of great benefit. I often receive data from other team members and it is clear that they have no knowledge of data management (i.e. storing multiple data sets on a single excel worksheet storing dates in text formats etc).

Social Sciences

- I have recently uploaded my publications to e-prints and realise that something similar would be very useful at a personal level for my own documents particularly scans and .pdfs of papers questionnaires etc. It might be that such software exists and would be good to have suggestions about how to manage data as an academic over a long working life multiple projects umpteen computers and storage devices and in an endlessly changing IT world. Thanks for doing this work.
- It would be good to find out from iSolutions what is available re data storage and how best to use existing facilities. I also need facilities how to share large datafiles with external researchers.

Michael Whitton, Kenji Takeda Aug 2011 mw2@soton.ac.uk

The Institutional Data Management Blueprint project is funded by JISC under the Managing Research Data Programme. Acknowledgements: JISC programme manager, Simon Hodson; University of Southampton project contributors http://www.southamptondata.org/about-us.html