DOI: https://doi.org/10.14324/000.learn.10

Case Study 9

Challenges and Opportunities for Research Data Management in the Arts, Humanities and Social Sciences: a practitioner's viewpoint

Author: Paul Ayris - Pro-Vice-Provost (UCL Library Services), Co-Chair of the LERU INFO Community (League of European Research Universities) & Adviser to the LIBER Board (Association of European Research Libraries)
Email: p.ayris@ucl.ac.uk

9.1. CONTEXT

This Case Study is based on conclusions drawn from a study of the challenges and opportunities for Research Data Management in the Arts, Humanities and Social Sciences in a research-intensive university – UCL (University College London).[1] The SCONUL strategic dataset (with data taken from HEAS – Higher Education Statistics Agency) for UCL records 6,286 FTE staff and 31,793 FTE students; the mean for research universities (defined as universities who are members of Research Libraries UK) was 2,661 staff FTE and 19,712 FTE students.

The survey was open to all UCL research staff and research students[2] and available online over 5 weeks in January and February 2016. The 67 questions dealt with respondents' awareness of policies and UCL services; with their practices of data management planning, data creation, storage and sharing. Finally, they were asked about their needs in terms of support and training. All questions addressed the respondents' most recent research project.

306 fully completed surveys were received (out of 619 unique surveys sent in). 130 research departments, institutes, centres and units were represented among the responses (out of a total of 380[3]) and were drawn from all UCL faculties.

The majority of responses came from research staff members, who are collaborating with other researchers on their project (either based within their department or external to UCL) and who have received external funding for it. The detailed responses from 3 of UCL's Schools and Faculties are given in the Appendix. These are the Faculties of Arts and Humanities, Laws, and Social and Historical Sciences.

9.2. ANALYSIS

9.2.1 Pan-UCL findings

Overall, the findings across UCL were as follows:

A very positive 70% of respondents are aware of UCL's and of their funder's policy on research data,

1 UCL Discovery: http://discovery.ucl.ac.uk/1540140/; last accessed 9 February 2017.
2 In this report, "research staff" encompass two categories of staff used by UCL Human Resources: "Academics" and "Researchers" (both full-time and part-time employees). It does not includes "Teachers". "Research students" refer to full-time Graduate Research students.
3 As listed in the UCL Departments A to Z (http://www.ucl.ac.uk/departments/a-z/, accessed 4 August 2016).

and 60% of respondents know about the UCL services related to Open Access. However, the level of awareness is problematic when it comes to internal research data-specific services: both the Research Data Management website (online since September 2015) and the Research Data Storage facility (available since 2012) are unknown to 60% of the participants.

The most common types of digital data created by respondents are spreadsheets, texts, databases and images. Remarkably perhaps, the answers also show that 30% of respondents produced non-digital data as part of their most recent projects and another 30% collected personal or sensitive data. Half of the respondents produced less than 100 GB of data over the lifetime of their project.

Data storage and archiving practices are also shown to be problematic. The most common method for storing research data was by using a personally-owned computer (45% of responses); the other favourite choices were a UCL computer, an external hard drive/USB stick or a cloud service. At the end of their project, half of respondents left their data on existing storage and, worryingly, 20% either could not recall exactly where they had archived their data, or had no plans for long-term preservation.

Among those who archived their data, 50% did it for their own re-use; for 20% of research staff it was because of funders' requirements. Half of the respondents have already shared their data with other researchers. Among them, only 25% did not have any concern when sharing data. When concerns were expressed, they were linked to legal questions, misinterpretation and time spent to collect the data.

A very large proportion of respondents (71%) said they thought about data management very early on in their projects, and a third indicated having someone in their team or department responsible for RDM. Yet, when asked what challenges they faced when managing their research data, a long list of problems enumerated by 217 participants is striking.

What is also surprising is that respondents mainly described challenges that are linked to handling data during their projects (storage, dealing with large volumes of data, good record keeping and backing-up procedures). This could indicate that they are not aware of where to find central information on these issues; or that the help available (whether at the central, faculty or department level) is not sufficiently adapted to assist with these essential measures.

Among the options proposed to them, respondents indicated that they would like help primarily in the following areas: storage and preservation of data; writing Data Management Plans; costing data management; data sharing and Open Access to publications. They would prefer to receive such assistance through online resources, training sessions in their department and regular drop-in sessions.

9.2.2 Faculty-specific findings

In terms of levels of awareness of policies and services, Arts, Humanities and Social Science researchers showed low levels of awareness of internal UCL RDM facilities and services. 66% of Arts and Humanities researchers did not know of the UCL RDM website. For Laws, half did not know. With regard to usage, only 10% of Arts and Humanities researchers and 8% of Social and Historical Sciences researchers had actually utilised it. This compares with, say, the Faculty of Engineering, where only 10% of researchers knew about the RDM website and 66% did not (Engineering had 67 respondents from 72 surveys).

As to the creation and analysis of data, an interesting picture emerges. For Arts and Humanities researchers, the most important type of data created was textual, with spreadsheets coming a close second. For Laws it was databases, with text coming a close second. In Social and Historical Sciences, the most popular forms were spreadsheets and photographs/digitised images, followed closely by databases and text. In Engineering, again the most popular formats were text, followed by spreadsheets and other images. Clearly, there is not the difference between the disciplines that might have been imagined.

Where there *is* a difference is in the size of the datasets created. For Arts and Humanities, only 3 respondents created datasets of 1-10 GB. The figures in Laws are too small to use to draw comparisons. In Social and Historical Sciences, the single biggest category for size of dataset creation was 1-10 GB. This contrasts with the Faculty of Engineering, where 11 of the 58 respondents were creating datasets of 1-10 GB, 12 datasets of 10-100 GB, and 11 datasets of 100 GB-1 TB.

When it came to storing and archiving research data, researchers in the Arts and Humanities commonly selected the hard drive of a personal PC or laptop, or an external hard drive/USB stick. For Laws, the most popular storage medium was an external hard drive/USB stick. In Social and Historical Sciences, the preferred media were the same as for Arts and Humanities.

For long term archiving, researchers in Arts and Humanities preferred subject repositories, or repositories external to UCL. In Laws and Social and Historical Sciences, researchers preferred existing storage. In the latter, many respondents admitted to having no archive plans. Engineers showed a pattern similar to Social and Historical Sciences. They preferred to use existing platforms for long-term storage.

Researchers were asked if they had any concerns about sharing their research data with others. Arts and Humanities on the whole had no concerns, and the same is true for Laws. Social and Historical Sciences did have concerns, however, and the most cited reasons were confidentiality issues/IPR or Data Protection.

Finally, researchers were asked what kinds of support they needed. In Arts and Humanities, researchers cited three main areas: storage and preservation of data, Open Access to publications and Data Management Plans. The same three preferences were cited in Social and Historical Sciences. In Laws, however, the most requested area for help was Open Access to publications, followed by a group who felt they needed no help. By way of illustration, in Engineering the most mentioned common areas for additional support were storage and preservation of data and Data Management Plans.

9.3. CONCLUSIONS AND RECOMMENDATIONS

A number of conclusions and recommendations can be drawn from the UCL survey, which illustrate the challenges and opportunities for research data management in the Arts, Humanities and Social Sciences:

Recommendation nº1

In all disciplines, research funders expect grant applicants and grant holders to explain how they will manage their data and to comply with their Data Management Plans. Being aware of these policies and services is a key element to writing successful funding applications. The earlier researchers receive assistance, the lesser the eventual risks for their projects.

- Faculties and research departments are encouraged to promote data storage solutions which comply with institutional and funder policies.
- Establish a central information service to support research data management

activities, such as http://www.ucl.ac.uk/library/research-support/research-data, and ensure this is continually promoted across the institution.

- Where possible, Heads of Departments should periodically invite the institutional Research Data Management team to give brief presentations to staff and research students on what assistance is available to them, including on 1-to-1 support and review of Data Management Plans.
- PhD students should be urged to attend courses on research support as part of an institutional Doctoral Skills Development Programme.[4]

Recommendation nº2

Training and support opportunities for both research staff and research students should not overlook the aspects around personal/sensitive data and databases as a large proportion of researchers use these as part of their projects.

Using personal computers and commercial cloud services to store research data represents a clear security risk for any data and a potential breach of security regulations if these are personal/sensitive data. An increasing number of funders currently expect that research data should be preserved for at least 10 years. Whether using an institutional facility or a discipline-specific repository, researchers should ensure that they know how to find reliable archiving facilities.

Recommendation n°3

The lack of clarity on where to find solutions to all of the challenges cited by research staff is a worrying observation. Academic faculties should be strongly encouraged to consider appointing/designating permanent staff members to assist researchers with data management in their subject disciplines.

Following these recommendations will help to avoid rushed and potentially costly short-term decisions; a lack of support when problems arise; and the outdating of skills and standards.

9.4. APPENDIX

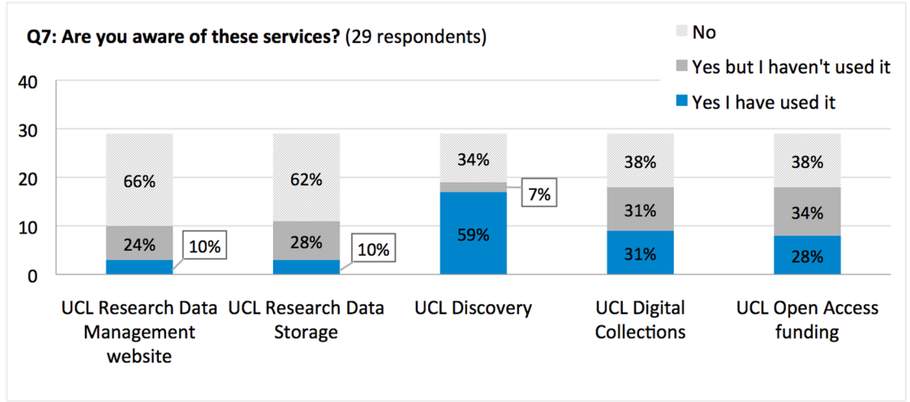
Selected detailed responses to the UCL Questionnaire

9.4.1 Faculty of Arts and Humanities

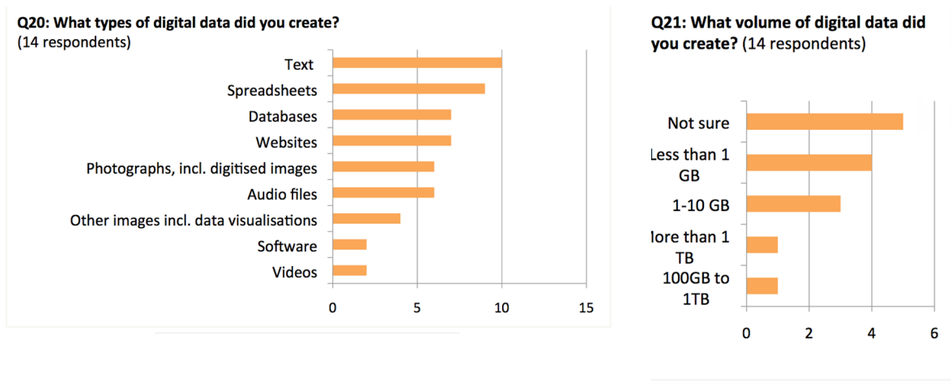
16 completed surveys were sent (out of 37 unique surveys transmitted) from researchers and research students in the Faculty. The overview below highlights responses to some of the key questions. It should be read after the Executive Summary and in conjunction with the whole report. The Research Data Management team is available to discuss the results.

[4] See, for example, http://courses.grad.ucl.ac.uk/; last accessed 29 January 2017.

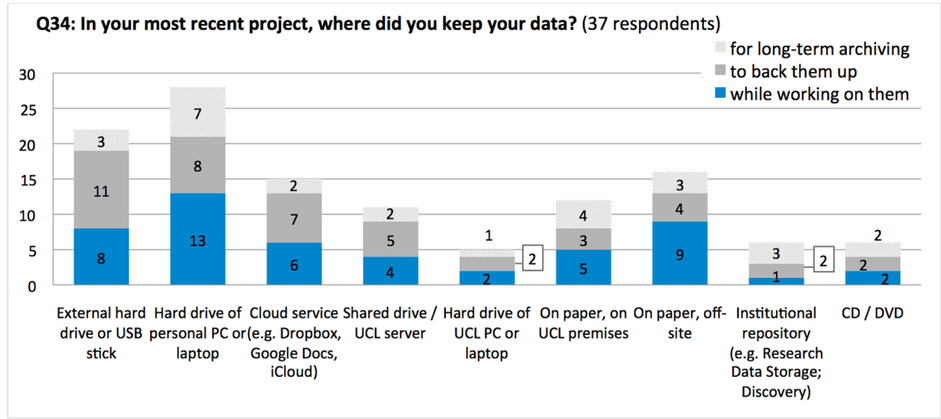
Awareness

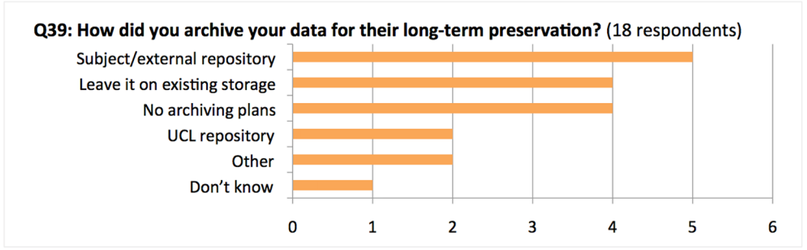


Creating & analysing data

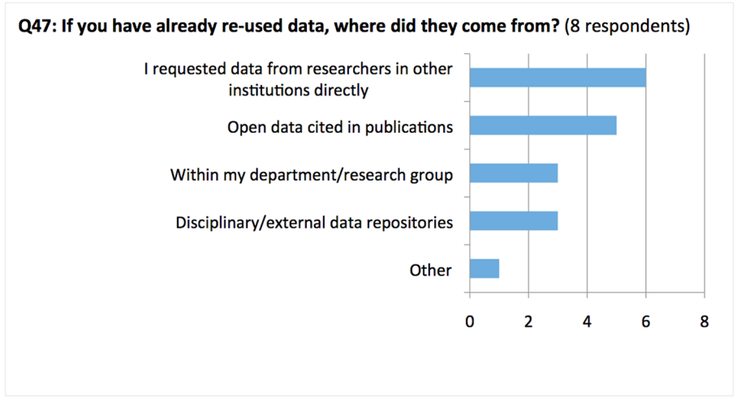


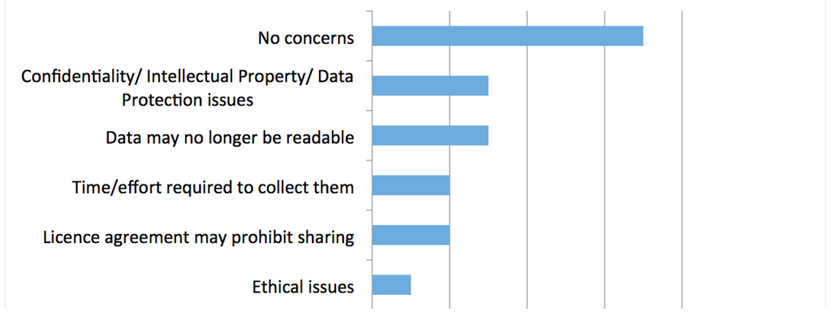
Storing & archiving data



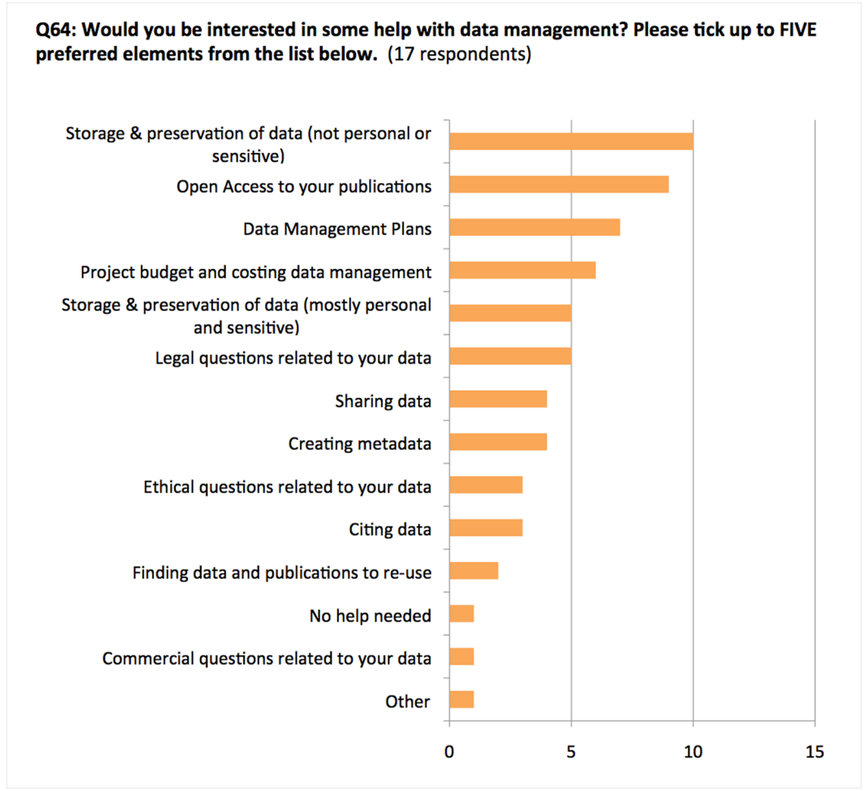


Re-using & sharing data





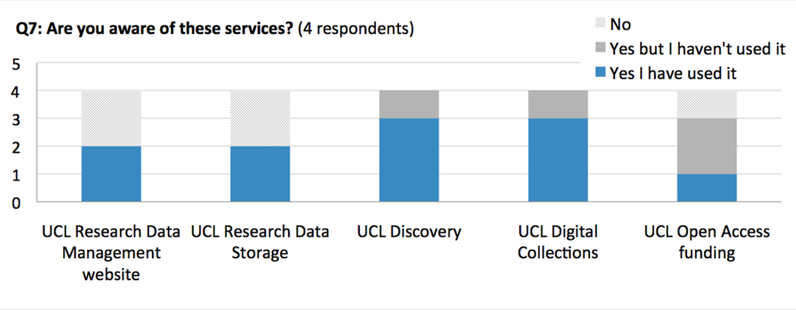
Support needed



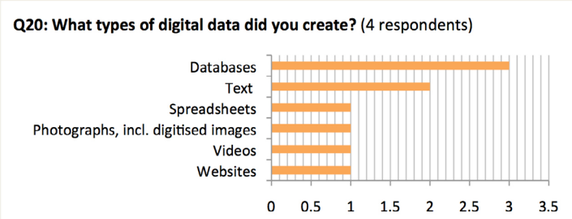
9.4.2 FACULTY OF LAWS

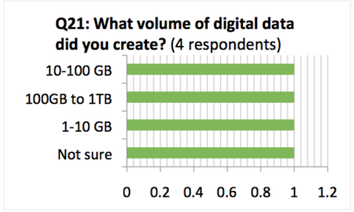
4 completed surveys were sent (out of 5 unique surveys transmitted) from researchers and research students in the Faculty. The overview below highlights responses to some of the key questions. It should be read after the Executive Summary and in conjunction with the whole report. The Research Data Management team is available to discuss the results.

Awareness: policies, UCL services & Data Management Plans

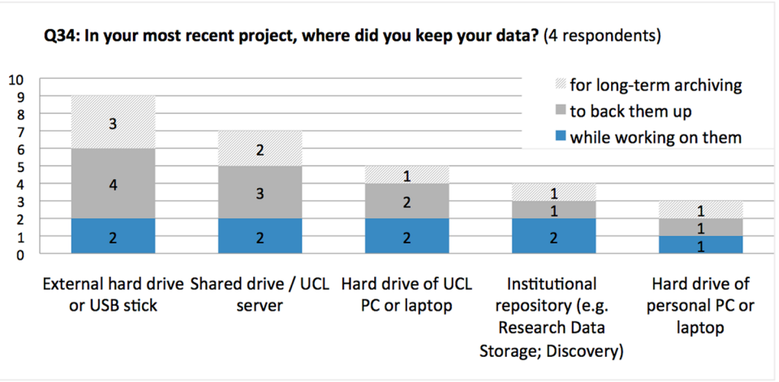


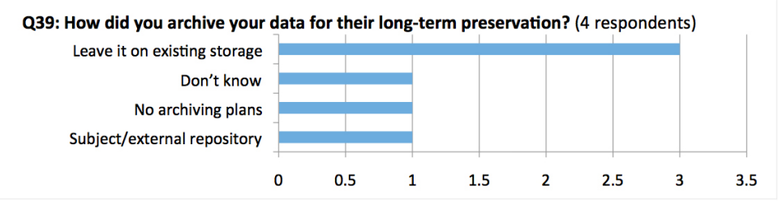
Creating & analysing data



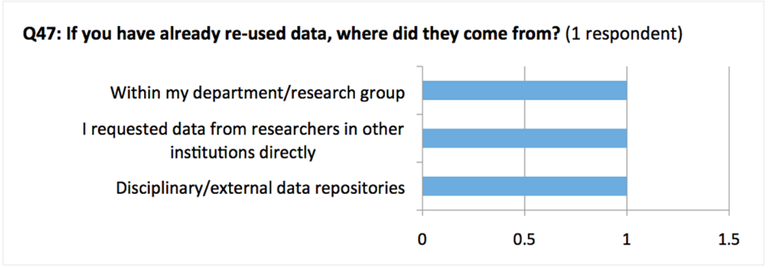


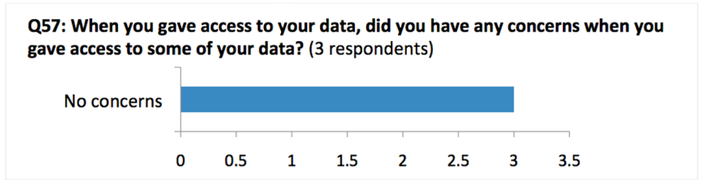
Storing & archiving data



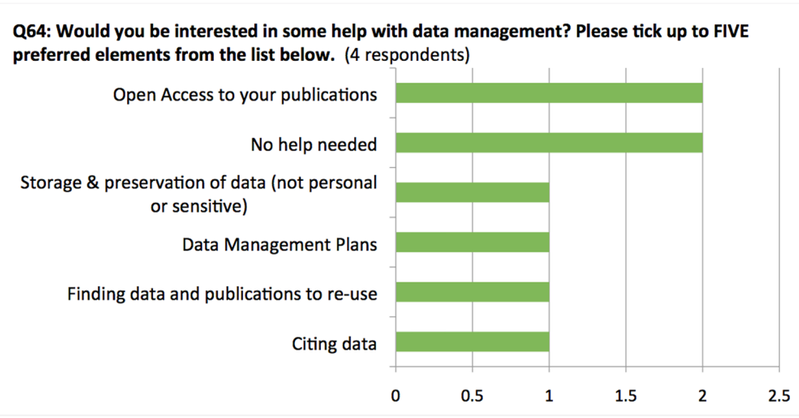


Re-using & sharing data





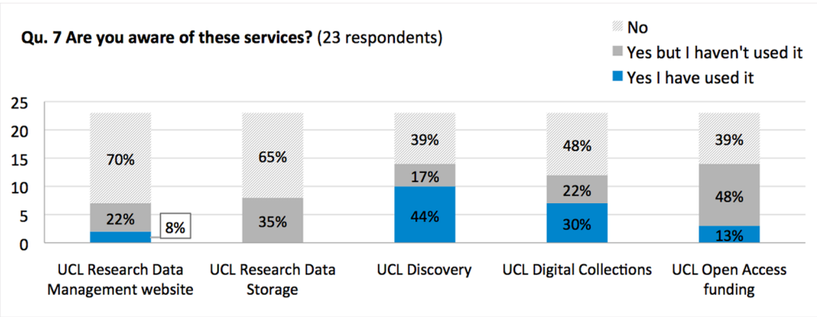
Support needed



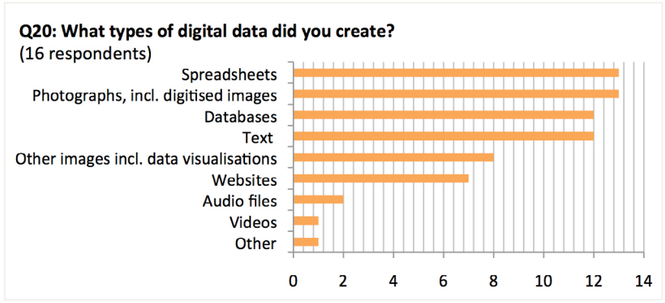
9.4.3 Faculty of Social & Historical Sciences

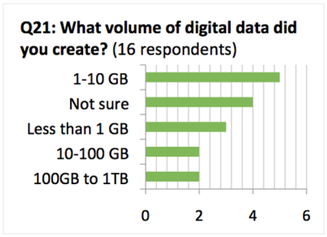
17 completed surveys were sent (out of 28 unique surveys transmitted) from researchers and research students in the Faculty. The overview below highlights responses to some of the key questions. It should be read after the Executive Summary and in conjunction with the whole report. The Research Data Management team is available to discuss the results.

Awareness: policies, UCL services & Data Management Plans

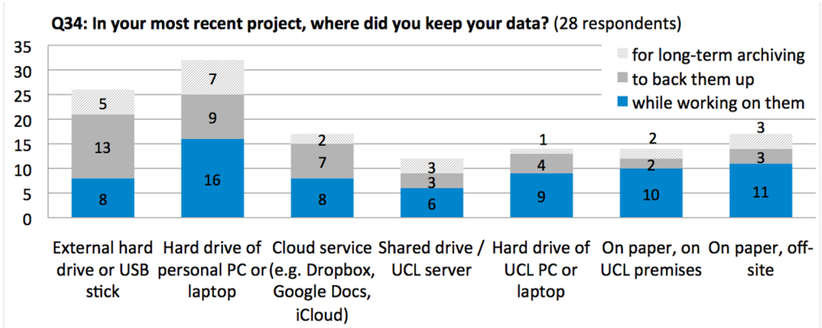


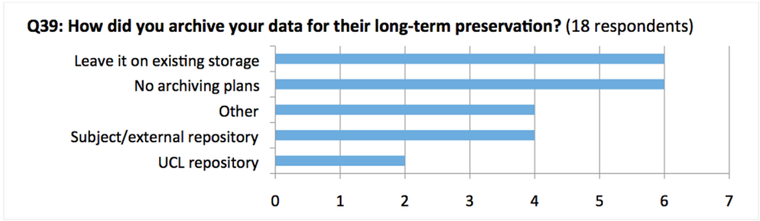
Creating & analysing data



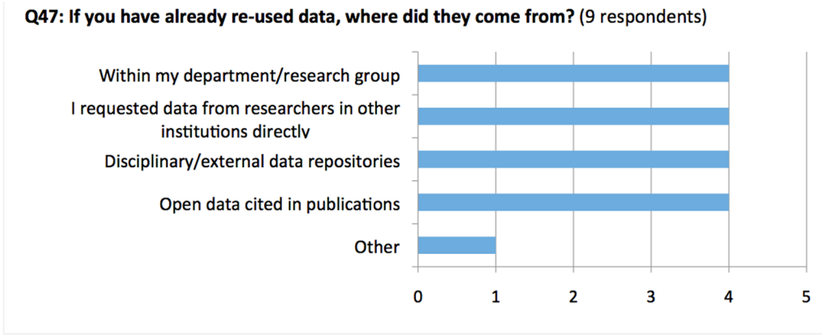


Storing & archiving data

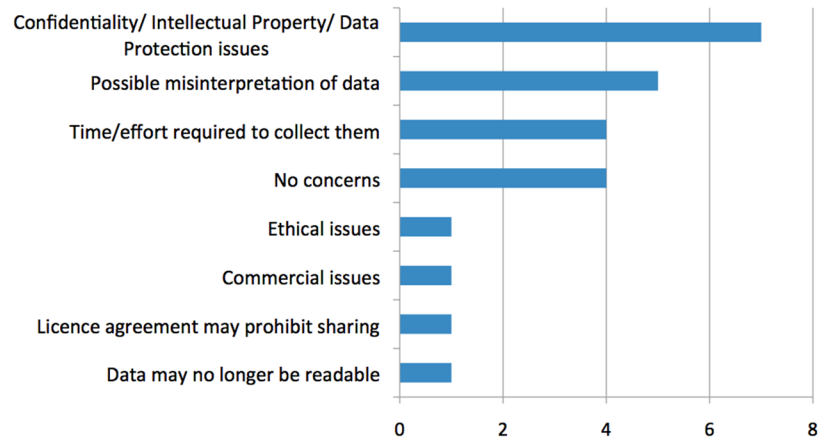




Re-using & sharing data



Q57: When you gave access to your data, did you have any concerns when you gave access to some of your data? (12 respondents)



Support needed

Q64: Would you be interested in some help with data management? Please tick up to FIVE preferred elements from the list below. (17 respondents)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.