**University of East London Survey on Research Data Management 2019**

***A snapshot of practice***

**Summary**

A Data Asset Framework survey was conducted by Library and Learning Services to learn more about opinions and practices in research data management at the University of East London. It was hoped that a deeper understanding would enable development of technical infrastructure and advisory services as appropriate, delivered by LLS, IT Services, or other relevant stakeholders.

Unfortunately, the survey received a low response rate due to several factors, including the time of year it was conducted. It has allowed us, though, to take a snapshot of data management practice at UEL—especially among postgraduate research students. UEL has a low number of externally funded researchers, meaning there are not many who are subject to the type of mandates and best practice for data management and sharing imposed by the Research Councils UK, for example.

Changes in practice have since emerged in response to the new UEL Research Data Management Policy[[1]](#footnote-2), published just a few weeks before this survey was launched: these are mainly represented by a dramatic increase in the number of data management plans being written by researchers (both staff and postgraduate). This has allowed us to identify trends and ‘pain points’ for researchers managing their data throughout the lifecycle, many of which reinforce the findings of the survey.

**Over 63% were collecting personal or sensitive data.**

**Nearly a third stored *all* their research data on an external drive or other portable media.**

**50% are using cloud services with personal, or combination of personal and institutional, accounts.**

**Researchers are using portable drives and cloud services for backing up their data.**

**A quarter of researchers said they would delete data of long-term value at the end of a project.**

The importance of data management planning cannot be overstated, in terms of benefits to researchers, but also as an opportunity for us to identify potential issues and recommend best practice where possible before data collection takes place.

Infrastructure for research data should provide robust and appropriate backup of data and secure, simple, file transfer to minimise use of non-institutionally provided methods.

There needs to be a better understanding of anonymisation and pseudoanonymisation among postgraduate researchers, and how these relate to GDPR compliance and storage of data.

Training on or signposting to resources, including best practice in data security, encryption, storage and back up methods available needs to be improved. Uptake of the services UEL currently offer could be increased if there was better awareness of what functions they provide and how, such as sharing via OneDrive.

Further advocacy and raised awareness of the research data management services and support provided at UEL, both technical and advisory, should also take place.

***Carly Lightfoot, Scholarly Communications Manager.***

***“Research data management should be a compulsory seminar series for PhD researchers”***

**Data Availability**

Dataset containing the full survey results, as well as a master list of survey questions, are openly-licensed and available at DOI: [*10.15123/uel.87551*](https://doi.org/10.15123/uel.87551)

**Background**

It was identified that current IT infrastructure for the handling of ‘active’ research data at UEL could be improved for the purposes of those conducting research at the institution. Following a meeting between Library and Learning Services and IT Services stakeholders, it was decided that more understanding of current practice by researchers was necessary to ensure our service offering met the demands of it. A survey was agreed as the best method to gain these insights. It was felt that the survey would also be a good opportunity to assess needs for other advisory and technical RDM services provided by LLS.

The populations we intended to survey were research-active staff members at the institution across all schools, institutes, and research groups, as well as postgraduate research students.

There were around 670 postgraduate research students at UEL and around 400 research staff at the time of the survey.

**Ethical approval**

Ethical approval was required for this research project and applied for in line with standard University Research Ethics Committee procedure. A Data Management Plan was written to accompany the ethics application, which was approved on 04.04.19.

**Methodology**

**Design and Dissemination of the Survey**

The survey was based on that developed for the Jisc Data Asset Framework Toolkit 2016[[2]](#footnote-3), which provided guidance for higher education institutions in implementing a Data Asset Framework (DAF) survey to collect information on what data they hold and what practices exist to store and share them at the institution. The Toolkit was part of Jisc’s research data shared service project (RDSS) and the original DAF was developed in 2009 by HATII and the Digital Curation Centre with funding from Jisc. It was chosen as the question list was believed to be able to elicit the type of information we intended to gather, plus we would be able to benchmark our findings against the other institutions who had carried out the survey and shared their data openly.

**Survey software**

The Jisc Online Surveys tool (formerly BOS: Bristol Online Survey) was chosen as the software to run the survey on. UEL has a subscription to the software, it supports conditional logic, and data is held within the European union.

**Survey questions**

Questions were selected from the Master question list of the DAF Toolkit 2016 survey. Using the master list will enable the shared dataset to be used for benchmarking among other HEIs who have carried out similar surveys as part of the RDSS project and for UEL again should we run another in future: it also provided a guide to which questions might be effective in gaining the information required of a successful Data Assessment Framework exercise. Questions are largely optional and some free text questions are included to allow respondents to expand on their views if unable to through pre-set questions. Personal information such as names and email addresses were not collected, but respondents were directed to the central RDM mailbox (researchdata@uel.ac.uk) should they wish to further discuss any issues.

The Master list comprises 54 questions, plus 6 additional questions related to software and code.

27 questions were selected for the UEL survey, plus another 5 which were designed into the survey as sub-questions rather than being standalone (e.g. if ‘yes’, then question). These were selected to suit the objectives of the survey and the institutional context. There were also some efficiencies made: disregarding Q15 *Do you apply any security measures to protect your data (e.g. password protection, encryption)?* and instead just using Q16 *Which of the following security measures do you use to protect your data?*, the answers to which include ‘none’ and ‘not sure’. It was hoped this would help response rates, as we managed to keep the time required to complete the survey down to approximately 10 minutes.

**Dissemination of the survey**

Several routes were identified for dissemination of the survey, including email via the networks of key stakeholders, the staff newsletter, intranet, and social media: it was hoped distributing through a mix of channels would maximise visibility of the project and help reach more of the target population.

Key stakeholders were identified as: PVC Impact & Innovation; Head of Graduate School; Data Protection Officer; REF Excellence Manager; Directors of Research; Academic Liaison Librarians.

A personalised survey link was used to give the project more credibility: <https://uel.onlinesurveys.ac.uk/rdm>

Responses were monitored while the survey was open and reminders sent at the halfway point to stakeholders as well as regular Twitter messages and another item in the ‘UEL In Focus’ newsletter.

**Limitations of the Study**

Unfortunately, the survey received a very low response rate, likely due to the time of year: it ran for 6 weeks, 24th April to 7th June 2019, which coincided with the end of the Easter holidays and during assessment & marking periods. This may have affected the staff/student ratio in respondents.

A lack of incentive may have also affected the number of responses: a financial reward or gift vouchers may have driven participation.

Due to selection bias there are areas of research that are under-represented, or not represented at all: indeed some researchers, for example in practice-based arts, may not have participated due to how data is perceived in those disciplines. Whilst this means we cannot draw firm conclusions across research activity, the results do at least offer a snapshot of activity and opinions around research data management at the institution.

22% of those visiting the survey completed their responses.

**Respondent Profile**

17 (81%) were Postgraduate research students (PGRs) and just 4 (19%) were members of research staff. 40% of respondents, including 2 members of staff, were affiliated with the School of Psychology. There were none from Architecture, Computing, and Engineering, or Arts and Digital Industries (now Arts and Creative Industries), therefore there are important areas of research that are not represented and should be focussed on in future information gathering.

Only 3 were funded by external agencies, significant due to organisations like the research councils often having clear mandates for researchers around managing research data, including storage, sharing, and archiving.



**Results of the Survey**

**Types of data**

Researchers are generating a variety of different types of digital data, but most commonly text-based files such as Word or PDF documents, and audio files like recorded interviews for example. The multiplicity of data types requires IT and RDM infrastructure to support the storage, analysis, sharing, and preservation of this data—not just standard files.

There are plenty of non-digital data being generated, mostly in the form of consent forms, notebooks/lab books, and questionnaires. There may be scope to digitise some of the paper-based data, or explore emerging tools such as electronic lab notebooks.





**Data Management Planning**

## Have a DMP: 13 (59.1%)

## Do not have a DMP: 4 (18.2%)

## Not sure: 5 (22.7%)

Since the publication of the new RDM policy, the writing of data management plans for research projects has increased dramatically. The library reviewed just 8 plans in 2018, this has risen to over 80 in 2019. Leading up to the policy publication, it was implemented in September 2018 that applications to the University Research Ethics Committee must be accompanied by a data management plan. It is expected that survey respondents who stated they had a DMP wrote them due to this requirement: 10 respondents cited for their reasons for having a DMP were that they are required by the institution, 10 that they were required by their supervisor/research group leader, and 10 that it was good research practice. Those that did not have a DMP said this was because they lacked knowledge or guidance, or due to the time and effort needed.

**Who needs access to your data?**

## Only myself: 17 (77.3%)

## Internal researchers: 6 (27.3%)

## External researchers: 3 (13.6%)

## Funders: 2 (9.1%)

## Publishers: 2 (9.1%)

Of those sharing data with external users, 3 were research staff and 1 a student: although there were a lack of staff completing the survey (4 in total), this implies that the need to transfer data outside of UEL may increase as researchers progress in their career. There are obvious implications in terms of the data transfer infrastructure UEL provides, especially with regard to data subject to the GDPR.

**Data sensitivity and security**

Over 63% of respondents collect some form of personal or sensitive data. If this is not managed properly, there is a heightened risk to the institution, and research participants, with regard to data security and non-compliance with the GDPR.

A variety of security measures were employed, most commonly password-protection, anonymisation, and physical security (locked rooms, for example). 41% encrypted their data, which is fewer than ideal given the nature of the data being processed. It is also worth noting that, in the experience of those reviewing data management plans submitted by researchers, there is some confusion as to the differences between anonymisation and pseudoanonymisation and the implications for GDPR compliance.





**Backup and storage**

## 45.5% had less than 50GB data

## 40.9% weren’t sure how much data they held

## 85.7% expected their storage requirements to stay the same or increase slightly or substantially over the next 5 years.

44% of respondents did not use university-managed network storage at all. Nearly a third stored all of their research data on an external drive or other portable media: only 8 of the 14 people storing data on portable media were using encryption to protect their data.



Cloud services are popular: just 3 researchers did not use one at all. The most used was OneDrive, likely as UEL provides access to the OneDrive for Business service. DropBox was also commonly used, with 44% storing data in this way. Half of respondents were using a personal account or a combination of personal and institutional accounts for cloud services. These storage options are not appropriate for long-term data sharing and preservation, and may not meet regulatory standards if personal data are being held there.



All respondents were backing up their data, mostly weekly, although 4 people did not have a regular backup schedule. Primary backup solutions were cloud services (40.9%) and external portable drives (36.4%).

**Data sharing and preservation**

A quarter of researchers claimed they would delete data of long-term value at the completion of a project and the majority, 45%, would leave the data in the same location.



**How do you share your data with others?**

## Email: 10 (50%)

## Cloud services: 6 (30%)

## Portable media: 4 (20%)

## Institutional file sharing service: 4 (20%)

There were mixed perspectives on sharing data publicly: only 1 currently sharing, 8 (36.4%) expect to in future, 6 (27.3%) unsure, 7 (31.8%) unwilling. However, researchers seem to be aware of the reasons or benefits of doing so.

There was a spread of reasons why respondents would not be willing to share their data publicly: 7 (33.3%) had not considered it—this may reflect the profile of respondent and their early career status, but makes clear that advocacy around data management and Open Scholarship needs to reach such researchers. 8 (38.1%) cited their data to be sensitive/confidential and only 2 said their data had no re-use potential.



**Knowledge gaps, concerns, training needs**

40.9% do not know what services are available at UEL to support data management and sharing, but 45.5% either use these services or expect to in future. There were a range of data management issues that researchers were concerned about: insufficient storage for just under half of researchers, insufficient security over confidential data for just less than half (but 23.8% extremely concerned), and 75% expressed some level of concern about their inability to interpret data due to, for example, poor or lost documentation or inadequate description.

In terms of training needs, it is clear that more provision needs to be made across the principles and practice of research data management. There were responses to all the suggested training subjects, but the most valued areas of training were long-term storage of data, data security, and publishing data.



1. University of East London, 2019, UEL Research Data Management Policy. DOI: 10.15123/PUB.8084 [↑](#footnote-ref-2)
2. Jisc Data Asset Framework Toolkit 2016, DOI: 10.5281/zenodo.177876 [↑](#footnote-ref-3)