Sending research material and data into the future

Malcolm Raggett
Project Officer
London School of Economics and Political Science
m.w.raggett@lse.ac.uk

Ed Fay
Project Manager
London School of Economics and Political Science
e.fay@lse.ac.uk

Why preserve research data?

There has been a sharpening of focus on open access to research outputs recently, with the publication of the Finch Report (Finch, 2012) then various supportive and non-supportive responses. Although the terms of reference for the Finch report included how to 'expand access to the quality-assured *published* outputs of research' [our emphasis] there is general acknowledgement amongst major funders that research *data* should also be made publicly available, free at the point of access, whenever possible. Unfortunately the use of the term 'data', which has come from the scientific community, suggests to many researchers in the social sciences, arts and humanities that it is only structured numerical information that is under consideration, so one of the first things we did was to adopt the phrase 'research material and data' to be more inclusive; to suggest that any information used as the basis for publication has the potential to be preserved and accessed by others. In this paper, though, we continue to use the single word 'data' on the understanding that it is in its broadest context.

Partnership between institution and researcher

There are several strands that need to be worked on in parallel in order to implement research data preservation:

- Since data preservation is a long-term commitment, just like a library or archive, the institution must buy into the process, so a high-level policy is needed that gives a thumbs-up from senior management. This policy is under development at LSE.
- The institution will need to develop and maintain an infrastructure to support long-term preservation. By infrastructure we mean the necessary human as well as the technological systems. The requirements of this infrastructure are currently being analysed at LSE.
- Support for researchers in terms of skills training, services that researchers could offload (such as file conversion) and awareness-raising. This is within the scope of the DICE project.
- Advocacy for the concept of data preservation from a culture-change perspective. This is also within the scope of the DICE project.

Current state of data preservation in higher education

It was recognised during the 1960s that digital data were being lost, and this caused the creation of several archives by UK Research Councils (<http://ukda40.data-archive.ac.uk/about/origins.asp>). These national archives have grown, merged and disappeared over the years but the need for their services has not diminished. If anything, the reverse is true: were it not for restrictive collection or curation policies, these national archives would be overwhelmed with data. What is done is done well, but there is a gap between their provision and a more general, albeit largely nascent, research data preservation requirement.

Current state of LSE researcher awareness and skills

We ran an online survey targeting all PhD students and research-active staff at LSE. Although we only achieved a 10% response (rather typical for this type of survey, unfortunately) the respondents were well distributed through the grades and across departments. This helped us get a picture of our researchers' skills and attitude towards data preservation.
When asked about their reasons for not making research data openly available, the main reasons were legal, ethical or reputational. Technological and cost factors were low in their consideration (Raggett, 2012. Question 17). Breaking down the results from this question further into PhD and non-PhD groups, we noticed that PhDs tend to be more cautious about making their data public, perhaps because they lack confidence early in their careers.

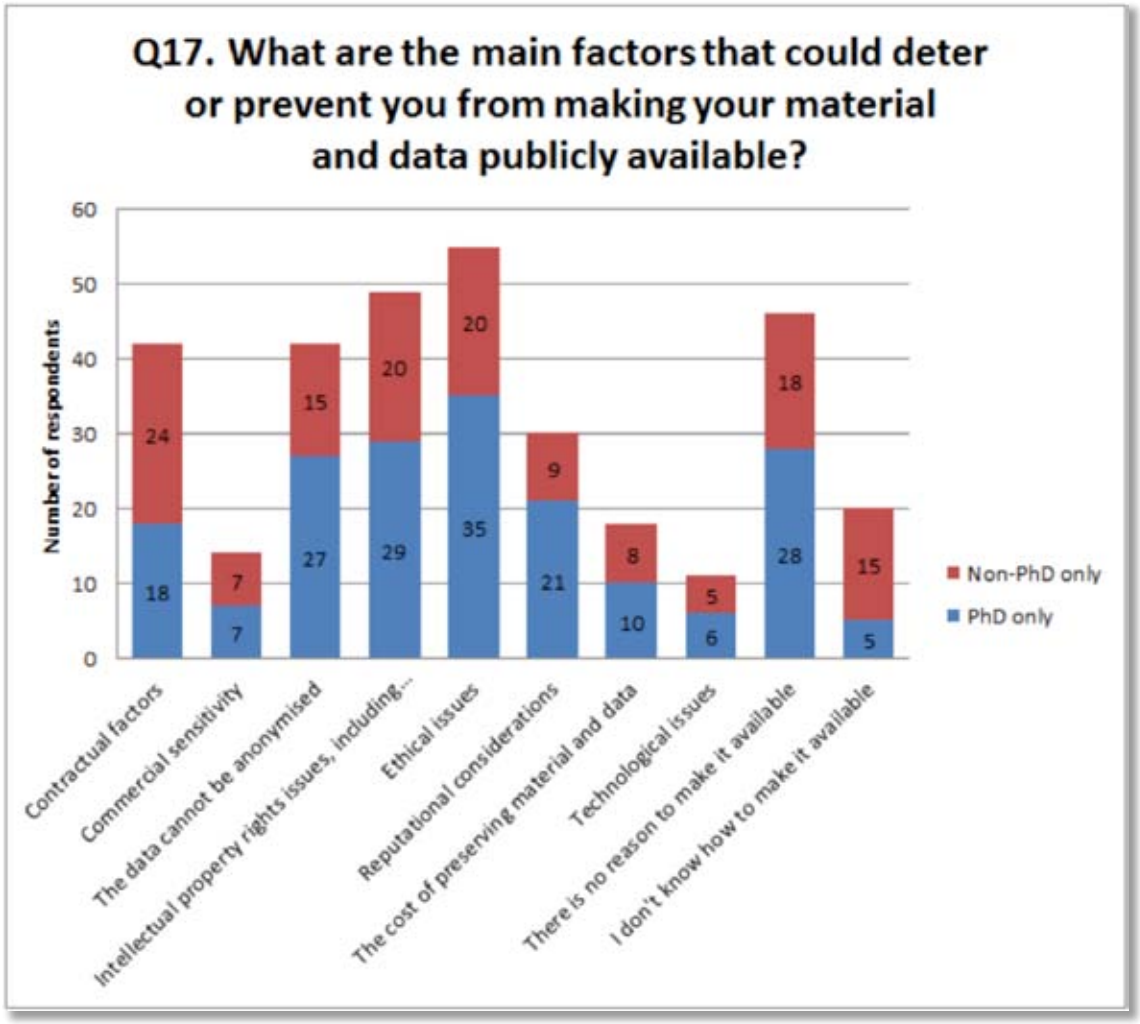

Fig. 1 Main deterrants/prevention factors

We asked about preferred delivery mechanisms for training and support and found that there is an aversion to most kinds of formal training including, interestingly, web-based training; the only exception to this is a 1-hour training course, which has a slight positive score. The most popular modes of delivery are informal and ad hoc training targeted at the researcher and their individual case. Rather surprisingly, the most popular of all the delivery mechanisms is a web-based FAQ approach. This is followed by email/telephone support, then discussions with a specialist. This tells us that it is vital that librarians are re-skilled to support researchers' data preservation needs and that we needed to provide an FAQ approach somewhere in our materials.

Our researchers' level of skill in information technology varied considerably from those who manage their data very competently down to those who only have one version of their data. We also noticed a willingness, particularly on the part of newer researchers, to use a variety of work methods that included internet-based editing, collaboration and storage. This is not something that has received encouragement from the School, and indicates a mismatch between demand and provision that needs attention beyond this project.

We found that many researchers disagree with the idea of open access to research data either actively (because they guard their data jealously) or passively (a 'nobody else would be interested in my stuff' or 'when am I going to get time to do that?' attitude). The benefits of making research data available were apparent to only a small minority of researchers, so there is clearly a need to be explicit about this. We also found that 90% of researchers don't know about or don't use any form of data management planning, which gives us scope for follow-on work.

Improving awareness and skill level

The implication for the DICE project was that we had to concentrate at two levels:

- Improving working practices throughout the research data lifecycle, not just at the point when 'preservation' takes place

- Encouraging a change in culture so that researchers accept the need to preserve their data and make it publicly available. We see this as the default position while recognising that there are legitimate reasons for not doing so.

And we had to do this for the researchers themselves as well as the staff in the library tasked with supporting the researchers.

Several other projects were funded by JISC in the same strand as DICE at the same time. Discussions with PrePARe at Cambridge and SHARD at University of London Computing Centre/I?//nstitute of Historical Research revealed similar outcomes from their user surveys and synergies between the general needs of the institutions' researchers. As a result, we collaborated on production of a leaflet and a generic list of FAQs. In both cases we were able to identify common messages that crossed subject domains and achieved a better product than if we had worked alone — and reduced the cost of printing too!

Project outputs

The training materials were always planned to be the major outcome from the project, though the literature review and user survey were vital intermediate steps. We took a tiered approach to structuring the training materials:

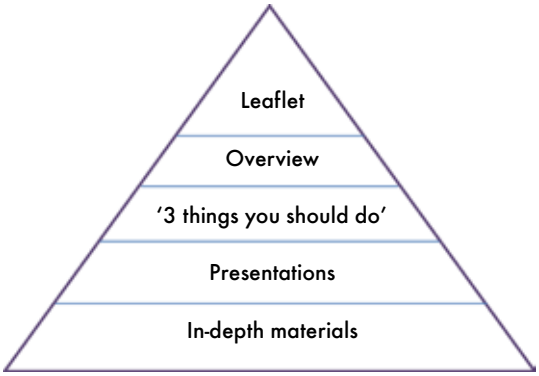

Fig. 2 Project outputs

1 An introductory leaflet that provided the outline for all of the materials: Explain it – store it safely – share it – start early
2 A more explanatory overview that could be presented to a group of researchers in less than 30 minutes
3 '3 things you should do' targeted at new researchers to encourage them into good habits from the start
4 Presentations summarising the in-depth materials
5 In-depth study materials for self-study or for trainers to adapt for their needs

The training materials and presentations are available from JORUM and learningresources.lse.ac.uk under a CC-BY-SA 3.0 licence. The literature review and user survey results are on the project blog http://lsedice.wordpress.com

The short duration of the project meant that we only had time to generate the training materials and test them briefly on library staff and a small cohort of PhD students; nevertheless, these initial tests indicated that we had developed them along the right lines and that they could be readily adapted to individual trainers' use.

Implications

Some projects are self-contained and their outputs final and static, but nothing could be further from the truth for the DICE project. Our outputs represent only the first step on what needs to be a sustained path of skill-building and culture-change for researchers and those who support them to send their research into the future. LSE library will be treating research data preservation as a theme in its forthcoming staff development programme; elements of the training materials will be incorporated into the course in information literacy to PhD students; and courses on research data preservation will be made available to all research-active staff and students during the next academic year.

Preservation of research data is only one aspect of the wider topic, research data management planning. Data management plans will be a requirement of all Research Council funding applications in the future, so LSE will need to develop a strategy for introducing this topic to our research community as the next step beyond data preservation.

References

Finch, J. 2012. Accessibility, sustainability, excellence: how to expand access to research publications. [Online] Available at: http://www.researchinfonet.org/wp-content/uploads/2012/06/Finch-Group-report-FINAL-VERSION.pdf
Accessed 1st October 2012

Raggett, M. 2012. Analysis of results from the research data survey. Full report. [Online] Available at: http://lsedice.files.wordpress.com/2012/04/survey-report-full.pdf
Accessed 1st October 2012