London School of Economics and Political Science
Digital Communication Enhancement (DICE) project

ANALYSIS OF RESULTS FROM RESEARCH DATA PRESERVATION SURVEY – SHORT REPORT

CONTENTS

1. BACKGROUND

1.1. The Digital Communication Enhancement (DICE) project has been funded by JISC to produce training materials to support digital research data preservation at LSE, and to make these materials available to the UK HE community. In order to focus the materials on the needs of the researchers, an online survey was conducted.

2. INTRODUCTION

2.1. A survey was conducted with the aim of discovering the awareness of and current practice relating to research data preservation at LSE. We targeted PhD research students and research-active staff by inviting these groups to respond to an online survey that was active from 14 to 25 March 2012.

3. SUMMARY

3.1. This questionnaire has shown the general lack of current awareness amongst LSE researchers of digital data preservation. It has also shown that there are cultural challenges to address as well as more technical training that is needed if researchers are to send their research data and materials into the future with confidence.

3.2. Researchers are using a wide range of technologies to create research data in a variety of formats, so as well as the general principles of digital preservation, we will use case studies of the major data types in the materials that DICE produces.

3.3. The survey shows that researchers have a preference for informal support and autodidactic means of learning. Although no one method is suited to everyone, we will target the methods of training and support that showed the most popularity, and use re-purposable materials wherever possible.

3.4. There is a presumption that data preservation and public availability of that data go hand-in-hand, however we identified a need for data preservation without public availability. There seems to be an acceptance by the researchers that this restricted preservation will take place under the control of the researchers themselves and on their own systems.

Nevertheless, training and support should be available to these researchers even if the School does not provide the repository for the data.

3.5. In general, research data are not well documented: researchers using others' data have reported difficulty understanding it, or even understanding their own data after a few years! The documentation of data is important when preserving it so this will be included in the DICE outputs.

3.6. Several issues emerged from the survey that are beyond the scope of the DICE project but which the School may want to consider. Specifically:

3.6.1. the experience profile of researchers at LSE shows a large number with less than 9 years' experience but then tapering rapidly to give a low number of more experienced researchers. The School may want to consider whether this balance is correct and whether its succession-planning for research continuity is adequate;

3.6.2. three-quarters of researchers recorded "self-funding" as a means of funding their research. The School may wish to investigate this statistic further;

3.6.3. 80% of LSE researchers do no data management planning. In an age of dependence on digital data this is a situation that needs addressing. A bid has been submitted to JISC for funding to extend the work of DICE, but alternative plans should be made if the bid is unsuccessful;

3.6.4. researchers are using cloud-based services as part of the tool set available to them for document maintenance, sharing and backup. LSE's policies and support for these services seem somewhat "behind the curve" compared with user behaviour and likely future demand. The School may want to re-examine its approach in this area;

3.6.5. some researchers, particularly those in the 6-12 year experience band, make use of departmental servers. The School might want to examine its costs and policy regarding the use of departmental servers versus the central network storage.

4. CONCLUSIONS AND RECOMMENDATIONS

4.1. This survey has shown that most LSE researchers have little awareness of the need for preservation of their research data, let alone the techniques and facilities to do so. This is no surprise as it is only recently that the desire to preserve research data, as opposed to publications, has arisen in most of LSE's disciplines.

4.2. The drive to preserve data has come principally from funding bodies such as the Research Councils, but this survey shows that only a small number of LSE researchers are funded from these sources. It is unlikely that the diverse governments, charities, endowments and even LSE scholarships, which provide a great deal of our researchers' funding, will introduce conditions that require them to preserve their data or make it publicly available in the near future, so any drive to raise awareness will need to be based on data preservation being for the common good or to the researcher's own advantage. Nevertheless, a link between the Research Councils' requirement for data planning, of which data preservation is one part, would be sensible to get maximum value from the training materials and to future-proof the material as much as possible.

4.3. Although many researchers plan to preserve their data beyond the life of the research project, most do not expect to make it publicly available but rather keep it restricted to some degree. In some cases this restriction will be the researcher's "default setting" and it should be possible to encourage a greater openness particularly with the PhD students, though it will take years for such a culture change to spread widely. In other cases the researcher will have valid justification for restricting the data, such as on ethical or legal grounds, in which case preservation without public availability will be needed. Such preservation will almost certainly be undertaken by the researchers themselves without reliance on

institutional or national repositories. This points us in the direction of personal archiving[1] as a potential source of techniques that we could harness for DICE, obviously adapting them for use in academic research.

4.4. Few researchers document their data to aid navigation or understanding of that data by themselves or others in the future. This is despite a number of researchers reporting problems when using other peoples' undocumented data. A significant habit that needs to be encouraged is appropriate documentation, and material on this must be included as part of DICE's output. Since there are some researchers who do document their data, particularly in the 6-9-year experience group, it may be possible to develop some case study material.

4.5. Researchers are using and creating diverse data types during their research. Text-based material, including emails, forms the bulk, but tabular material, charts and graphs are very popular, and computer code & scripts are also significant. These media will need to be prioritised, perhaps by means of case studies, in the DICE project outputs.

4.6. Some researchers are using Cloud-based services to develop, maintain and backup their data. DICE should include material to support these services, including advice on the risks involved.

4.7. Most, though not all, researchers are organising and backing up their files to some extent. Since both practices have a bearing on data preservation, advice on these topics will be included in the DICE outputs.

4.8. There is no method of training or support that suits everyone: we will need to aim for diversity of delivery using the most popular methods identified in the survey. These are:
4.8.1. Web-based FAQs
4.8.2. Email & telephone support
4.8.3. Discussion with a data preservation specialist
4.8.4. 1-hour group training session

4.9. Some support materials would also be useful in Moodle; these could be linked to the 1-hour training course but able to function separately too. Although DICE is specifically charged with producing materials for the MY592 Information Literacy course, and will do so, the evidence is that this will not be the most effective method of supporting data preservation. However, where possible the training materials will be capable of being re-packaged so that, for example, Moodle material can support the 1-hour training course and can be re-packaged for MY592. It is also clear that someone will need to maintain the list of FAQs as well as provide email support and specialist discussion. Learning material will be needed that should be able to facilitate the training of this person/team by providing more depth and detail than the training aimed at researchers. Moodle modules would be ideal for this, and could also be made available to researchers for personal study.

4.10. No strong synergies were found between research data preservation and administrative data preservation (Records Management). The training materials produced by the project will be freely available to re-purpose for Records Management, but they will not be specifically tailored for this.

---

[1] Personal archiving has developed largely in response to the increased interest in family history and the recognition of the potential value of "that stuff in the attic". Some archives, libraries and museums run courses for the general public in this topic.

Malcolm Raggett, 2012