Understanding Capacity in the Heritage Science Sector:
A response from the Digital Preservation Coalition

Introduction

1. The **Digital Preservation Coalition (DPC)** is a not-for profit membership organisation whose primary objective is to raise awareness of the importance of the preservation of digital material and the attendant strategic, cultural and technological issues. Its vision is to make **our digital memory accessible tomorrow.** This is our third contribution to the National Heritage Science Strategy, and is a response to Report 3: *Understanding capacity in the heritage science sector*, September 2009.

2. In this response we note that a lack of capacity imperils our digital heritage and constrains preservation planning in museums archives and libraries. **We call on the NHSS to address the lack of capacity for digital preservation in the UK.**

3. In our first submission we argued that digital resources can have *'lasting value and significance, and therefore constitute a heritage that should be protected and preserved'* (after UNESCO 2003) and assessed the distinctive challenges which digital resource face. In our second submission we agreed that there is a need for *'improved storage and long-term curation of digital data resulting from heritage science'* and listed relevant initiatives.

4. In this response we note our concern that **the distinctive capacity required to preserve digital objects has been overlooked by NHSS**. This is surprising because techniques appropriate to the preservation of digital objects are within the scope of the enquiry. Digital resources are a new challenge. They continue to grow in scale, pervasiveness and complexity and they impact directly with contemporary collecting in such diverse fields as modern art, technology and social history as well as research, librarianship and archives.

5. We present evidence of a critical lack of capacity. Typically this takes two forms: a lack of digital preservation skills amongst mainstream preservation practitioners; and a fragmented provision of services and tools. This is compounded by the gap between emerging tools and their intended users.

6. We note that the UK is well placed to resolve the challenges of digital preservation, largely thanks to a small group of researchers and developers who have taken a lead on the topic. Therefore **we call on the NHSS to support wider and more effective dissemination of skills, tools and services that will safeguard the UK's distinctive digital heritage.**

7. The Digital Preservation Coalition is a cross-sectoral organisation with a distinctive capacity to explore, develop and deliver co-ordinated approaches to the management of the UK's digital legacy. **We offer our support** in articulating this element of the National Heritage Science Strategy.

The skills gap: capacity for digital preservation in the UK

8. NHSS examines capacity by analysing the distribution of specialists within disciplines. However, it overlooks our capacity to sustain digital objects through time and it neglects to evaluate how effectively research in digital preservation is disseminated.

9. It has not been possible to furnish NHSS with a parallel analysis this sector. However, a number of recent reports demonstrate a significant skills gap for digital preservation in the UK. We encourage the NHSS to consider this evidence and act accordingly.

10. In 2006, the DPC evaluated the obstacles and incentives for continuing access to the UK's distinctive and growing digital estate (Sharpe and Waller 2006). Based on an extensive survey it delineated an objective, strategic plan through which the UK could tackle the challenge ongoing access to digital objects. 89% of respondents to the survey sought improved provision of guidance and training (Sharpe and Waller 2006, 31). Consequently, the report recommended 'provision of continual professional development for existing individuals with relevant skills' and that training in digital preservation 'should form part of the professional training for conservators, archivists and librarians.'

11. More recently, the DPC has surveyed local government archive services in England and Wales (Boyle et al 2008). All respondents listed skills and training as gaps to be filled (Boyle et al 2008, 5). 'Skills' and 'support' were listed as joint second behind 'funding' as obstacles to preservation (Boyle et al 2008, 4). One might infer that if 'funding' were available it would be spent on acquiring better support and more skills.

12. Comparison with other European nations shows that the UK is not alone. For example a recent survey in the Netherlands has concluded that knowledge and expertise of digital preservation is 'scattered and inaccessible' outside national agencies (Angevaare 2009, 17).

13. A pan-European survey by PARSE.Insight asked data managers within research institutions what they need in order to provide long term access to data. 'Expertise' and 'training' were only slightly less important than 'resources', and more important than 'more repositories' (van der Hoeven 2009, 16).

The skills gap: recognised but not yet addressed

14. The requirement for greater capacity for digital preservation has been echoed in previous policy consultations. It would be appropriate for NHSS to evaluate whether these policy resolutions have been delivered, and how they might align with them.

15. For example, the Office for Science and Innovation called for an 'E-infrastructure for Science and Innovation', which enables 'the preservation of digital information output as a vital part of the nation's cultural and intellectual heritage' (OSI 2007: 9). The emergence of such an infrastructure and allied expertise would help build capacity.

16. More recently, a consultation on 'Archives for the 21ˢᵗ Century' called for a 'co-ordinated response to the growing challenge of managing digital information' (HM Government 2009). It is too early to know how this might be manifested but there is an opportunity for NHSS to align itself with this initiative.

Filling the skills gap: current efforts incommensurate with the challenge

17. A small number of training providers have already designed and delivered training to help institutions deal with the mounting challenge of digital preservation. These should be taken account of by the NHSS in its discussions of how to fill these gaps. Though important, these activities are not yet commensurate with the challenge described. Therefore we call on the NHSS steering committee to consider how such efforts can be extended and sustained.

18. In 2005, the University of London Computer Centre, with assistance from partners developed the Digital Preservation Training Programme. This three-day course has been offered on 8 successive occasions with and the most recent course (October 2009) ran at full capacity with 22 attendees. Evaluation has been consistently positive and some of the early participants are now leading preservation strategies within key national institutions. However, with the exception of a small number of grants provided by the DPC, the course receives no external support. Attendance costs £650 per participant.

19. In 2009, the DPC, The National Archives (TNA) and the Society of Archivists (SoA) developed the 'Digital Preservation Roadshow'. This one-day training event will be repeated in 7 locations in Britain and Ireland until early 2010. Aimed primarily at the archivists it has been very popular, attracting audiences of between 30 and 50. It is subsidised by DPC, SoA and TNA. Members of the DPC or SoA can attend for free while others are charged £75.

20. In 2008, with funding from JISC, the Digital Curation Centre developed a three-day course entitled 'DCC Digital Curation 101', an advanced introduction to lifecycle planning of data management for bench scientists and researchers mandated to provide continuing access to data from research projects. The course has now run on 3 occasions (including one 'DCC Digital Curation 101 Lite'). This course is offered free of charge to participants though access is limited to higher education and cognate research institutions.

21. The DPC offers an annual programme of events themed around its Technology Watch Reports. These provide authoritative and concise guidance to the challenges of long-term access. Although popular and in spite of a long list of topics, the DPC is currently only able to produce three reports per year.

22. In addition the DPC supports attendance at training and workshops for members (such as the Digital Preservation Training Programme) through a Leadership Programme.

23. A range of projects and institutions offer focussed training in digital preservation. Although welcome and well-received, such ad hoc provision is unlikely to meet a wider strategic need.

The skills gap: conclusions

24. The NHSS has examined such significant subjects as routes into the heritage science sector (3.4), succession planning age profiles of practitioners (3.1.4) and access to resources. Without a substantial ancillary assessment it is hard to provide a parallel commentary on capacity within the digital preservation community. Nonetheless, baseline evidence indicates the poor distribution of digital preservation skills in the UK.

25. We call on NHSS to examine how these gaps can be filled and we welcome a dialogue on how to resolve pressing issues of capacity for the preservation of our digital heritage.

Tools and services for digital preservation

26. In the following section we highlight gaps in tools and research. The sector is increasingly well served with tools and there is an increasing volume of emerging research. This is not to say that solutions exist to every digital preservation challenge: but it does support the optimistic conclusion that solutions are possible. However, the previous section has shown that the gap between the small number of tool-developers and the wide number of potential users is lamentably wide. **Therefore we call on NHSS to support more effective dissemination of digital preservation tools and services.**

27. This rapid development has transformed digital preservation in the last decade and the UK is well placed to face the challenge of digital preservation. Consequently, it is no longer possible for heritage managers to abdicate responsibility for long term access to digital resources on the premise that it is unachievable.

28. For example, The National Archives have developed and released a number of core tools which support file format management. An extensive format registry called PRONOM lists the characteristics of known formats, while a freely-available application called DROID can characterise and list files across a network.

29. The Digital Curation Centre, with funding from JISC has developed a range of assessment tools for digital preservation such as Digital Repository Audit Methodology Based on Risk Assessment (DRAMBORA) and the Data Audit Framework. These support data managers by helping to scope requirements and improve practice through ongoing review.

30. The PLANETS partnership – led by the British Library – has been funded by the European Commission to develop and provide a range of tools for the digital preservation community. Due for completion in 2010, tools like PLATO provide an integrated preservation planning framework. Components are designed to be modular and interoperable permitting flexibility, thus permitting integration into a variety of preservation work flows and services.

31. These tools have made a significant contribution to the preservation community, but they are not static. For example, The National Archives has recently completed a user evaluation

in advance of a 5th release of the DROID tools. Similarly the PLANETS project is evaluating business plans for maintenance of its tools once capital funding is exhausted. Digital preservation tools are dynamic and need to remain dynamic.

32. The digital estate continues to grow in scale and complexity so a challenge for these tools is the extent they can cope with this expansion. For example, research shows metadata creation is a significant cost driver for digital preservation (Beagrie et al 2008). Tools that support automated metadata extraction are likely to become critical as the number and complexity of files increase.

33. Therefore, in order to support ongoing maintenance of digital preservation tools the heritage science sector will need to develop and maintain a close relationship with elements of computational sciences. For example, linguistic and statistical analyses are active areas of research which can enhance and support the development of such tools.

34. This relationship with aspects of computer science also brings significant advantages for the creation, analysis and reuse of data within conventional heritage sciences. For example, emerging tools for linguistic analysis or statistical modelling of data; solid modelling within architecture; and geo-spatial analyses within the historic environment will enhance our understanding of the physical heritage. Tools to link, share and describe the capabilities of data have seen active development over the last decade and have been applied to functions of heritage management at local, national and international scales.

Preserving 'Grey Literature' and other research outputs

35. We welcome the attention paid to the preservation of 'grey literature' that emanates from research in heritage science. However, processes of heritage science produce considerably more digital resources than simply grey literature. Assuming that services can be developed to preserve and disseminate grey literature, then underlying data sets should also be made available. For example, the NHSS pays particular attention to staff involved in digitisation and scanning. Appropriate steps should also be made to preserve and disseminate data from such processes.

36. We welcome the suggestion that the ADS's OASIS grey literature service could be a model for the heritage science sector's efforts to disseminate grey literature and related data.

References

Angevaare, I, 2009 A Future for Our Digital Memory: Permanent Access to Information in the Netherlands (Interim Report, English Summary), NCDD, The Hague online at: http://www.ncdd.nl/en/documents/Englishsummary.pdf last access 22/10/09

Beagrie, N Chruszcz, J and Lavoie, B 2008 Keeping Research Data Safe: A Cost Model and Guidance for UK Universities, JISC, London online at:

http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf
last access 28/08/09

Boyle, F, Eveleigh, A and Needham, H 2008 Report on the Survey Regarding Digital
Preservation in Local Authority Archive Services, Digital Preservation Coalition, York, online
at http://www.dpconline.org/docs/reports/digpressurvey08.pdf last access 28/08/09

HM Government, 2009 Archives for the 21st Century - England: Consultation Draft online at
http://www.mla.gov.uk/what/strategies/~/media/Files/pdf/2009/archives-for-the-21st-
century-england

Office for Science and Innovation 2007 Developing the UK's e-Infrastructure for Science and
Innovation, online at http://www.nesc.ac.uk/documents/OSI/report.pdf last access
18/09/09

UNESCO 2003, Charter on the Preservation of the Digital Heritage, UNESCO, Paris, online at
http://portal.unesco.org/en/ev.php-
URL_ID=17721&URL_DO=DO_PRINTPAGE&URL_SECTION=201.html last access 04/08/09

van der Hoeven, J 2009 First Insights into Digital Preservation of Research Output in Europe,
PARSE.Insight, online at: http://www.parse-insight.eu/downloads/PARSE-Insight_D3-
5_InterimInsightReport_final.pdf last access 23/10/09

Waller, M and Sharpe, R 2006 Mind the Gap: assessing digital preservation needs in the UK,
Digital Preservation Coalition, York, online at:
http://www.dpconline.org/graphics/reports/mindthegap.html last access 28/08/09

<div align="right">

Dr William Kilbride FSA,
Executive Director of the Digital Preservation Coalition
October 2009

</div>