

So where's the black hole in our collective memory? A Provocative Position Paper (PPP)

[This is one in a series of thought provoking and controversial position papers on a range of issues surrounding digital preservation. It is our intention that these papers will promote vigorous debate within the digital preservation community and encourage people to think about digital preservation in new and innovative ways by exploring and challenging the received wisdom. These papers in do not necessarily represent the views of DigitalPreservationEurope nor our recommendations on the subject discussed.]

Has the digital preservation community cried wolf too often? Are its strident, alarmist proclamations about loss of digital materials too extreme?

The title of this paper is based on a quote from *Your Data At Risk* published by the National Council on Archives in 2005: 'If we do not manage our digital assets well, we are heading for a black hole in our and our organisation's and our country's collective memory' (National Council on Archives 2005, Section 5.2). Similar alarmist predictions are easy to find: 'there will be 'a digital black hole ... truly a digital dark age from which information may never reappear', note Deegan and Tanner (2002) and Jonas Palm titles his 2006 paper 'The Digital Black Hole'.

Similarly, the term *digital dark age* is frequently used to describe the issues that are faced by the loss of digital material. This term has even made it into the collective consciousness through a Wikipedia entry where 'the period around the turn of the 21st century [is] ... comparable to the Dark Ages during the Middle Ages in the sense that there will be a relative lack of written record' (Wikipedia, 'Digital Dark Age'). Again, it is easy to locate other uses of this term: for example, in *The Guardian* (Scholefield 2003), and in Dave Mattison's *Ten Thousand Year Blog* (Mattison 2004).

A *black hole* in our collective memory, *digital dark ages* – these are emotive terms. Is the situation they describe correct? Where's the evidence? Is the digital preservation problem really so great that we need to spend hundreds of millions of euros, pounds and dollars investigating and developing solutions?

Verifiable evidence is lacking. There is no equivalent for digital materials of Lost Memory, the UNESCO Memory of

the World's list of libraries and archives destroyed in the twentieth century (Hoeven and Albada 1996).

For digital materials, the English-language literature consists of either general statements about categories of lost material (for example U.S. federal government web sites) or a small number of frequently-reiterated specific examples (such as the BBC's Domesday Project, NASA data, the Viking Mars mission, the Combat Area Casualty file containing prisoner of war and missing in action information for the Vietnam War, the first e-mail, the first web site).

Many of the documented examples are in fact not about data loss, but about data recovery. Examples of this include the Challenger Space Shuttle Tapes, and German unification and the recovery of electronic records from East Germany (Ross and Gow, 1999, pp.39-42). The examples from the 1996 Task Force on Archiving of Digital Information have been reported widely. Its report notes, as one example, the case of the US Census of 1960. By the middle of the 1970s the Census Bureau faced the danger of losing this data, which was stored on obsolete UNIVAC type II-A tapes. But it was not lost, as by 1979 nearly all of the data determined to be of long-term value had been successfully copied onto industry-standard tapes. Three case studies presented in the Digital Preservation Coalition's 2006 report *Mind the Gap* are also about data recovery, not data loss (although this report does indicate that considerable expense was required to recover the data). The Viking Lander data from NASA's Mars probe in 1975 was recovered by re-inputting it from printouts. The 1986 Domesday Project was produced on laserdiscs viewed with software running on BBC Microcomputers and became inaccessible, but has since been recovered. The Schools Census data was compromised by the loss of metadata, but has since been recovered.



From the literature it is only possible to conclude that the evidence of digital data loss is, overall, anecdotal, and that data can be recovered, albeit at considerable expense.

The digital preservation community's inability to bring firm evidence to bear in support of its contentions about data loss, coupled with the alarmist rhetoric of terms such as *digital dark ages* and *digital black hole*, leave us exposed. Our efforts and our calls for resourcing can readily be ignored. We need to document more examples of data loss to supplement the largely anecdotal examples that are commonly provided. Some examples are probably present in unpublished reports or business cases, but the publicly-accessible literature contains few of these.

Further, the evidence that is currently available is unconvincing. It does not meet basic criteria for good research, such as building on what has already been investigated, usually presented as a literature review. Applying only this single criterion to the preservation of digital information would at the least call for a thorough literature search to ascertain, as accurately as possible, just how large the problem is. We need answers to three questions:

- 1) How much digital information has been lost and how much has been compromised?
- 2) To what extent has the data been compromised?
- 3) Is the problem of digital preservation as great as we have assumed?

Finding answers to these questions is crucial to securing resources to ensure the future of digital preservation. Inability to answer these questions will lead to scepticism about whether the problem is as great as claimed, and to lack of appropriate and adequate resourcing. This is an excellent reason to put some effort now into quantifying the extent of digital information loss or compromise, or, at the very least, to document more examples to supplement the few specific studies currently available.

References

- Deegan, M. and Tanner, S. (2002) 'The Digital Dark Ages', Update, May
www.cilip.org.uk/publications/updatesmagazine/archive/archive2002/may/update0205b.htm
(viewed 16 December 2007).
- Digital Preservation Coalition (2006) 'Mind the Gap: Assessing Digital Preservation Needs in the UK'
<http://www.dpconline.org/graphics/reports/mindthegap.html>
(viewed 16 December 2007).
- Hoeven, H. van der and Albada, J. van (1996) 'Lost Memory: Libraries and Archives Destroyed in the Twentieth Century' (Paris: IFLA and ICA)
http://portal.unesco.org/ci/en/ev.php-URL_ID=2165&URL_DO=DO_TOPIC&URL_SECTION=201.html
(viewed 16 December 2007).
- Mattison, Dave (2004) 'The Black Hole of Digital Archives', Ten Thousand Year Blog post 10 November 2004
<http://www.davidmattison.ca/wordpress/?p=851>
(viewed 16 December 2007).
- National Council on Archives (2005) 'Your Data At Risk: Why You Should be Worried about Preserving Electronic Records'
<http://www.ncaonline.org.uk/materials/yourdataatrisk.pdf>
(viewed 16 December 2007).
- Palm, Jonas (2006) 'The Digital Black Hole' Training for Audiovisual Preservation in Europe
www.tape-online.net/docs/Palm_Black_Hole.pdf
(viewed 16 December 2007).
- Ross, S. and Gow, A. (1999) 'Digital Archaeology: Rescuing Neglected and Damaged Data Resources: a JISC/NPO Study Within Electronic Libraries (eLib) Programme on the Preservation of Electronic Materials' (London: Library Information Technology Centre)
www.ukoln.ac.uk/services/elib/papers/supporting/pdf/p2.pdf
(viewed 16 December 2007).
- Schofield, Jack (2003) 'Digital Dark Age Looms', The Guardian 9 January 2003.
- Task Force on Archiving of Digital Information (1996) 'Preserving Digital Information: Report of the Task Force on Archiving of Digital Information' (Washington DC: Commission on Preservation and Access).