



ALIA 2006 Biennial Conference



Australian Library and
Information Association

Virtual Refereed Paper

Dr Ross Harvey

Charles Sturt University

Contact details

Postal: School of Information Studies
Charles Sturt University
Locked Bag 675
Wagga Wagga, NSW 2678

Phone: 61 2 6933 2369

Email: rossharvey@csu.edu.au

Biography

Dr Ross Harvey is Professor of Library and Information Management at Charles Sturt University's School of Information Studies. He has held academic positions in Australia, Singapore, and New Zealand, and has been a Visiting Professor at the University of California Los Angeles. He has published widely in the fields of preservation of library and archival material, bibliographic organisation, and library education, a recent book being *Preserving Digital Information* (München: K.G. Saur, 2005). His current research interests include the preservation of library and archival material, especially in digital form. Harvey was based at the National Library of Australia from April to June 2003 as a National Library Fellow, where he researched issues associated with the preservation of digital information, and he taught at the European Union's DELOS Digital Preservation Summer Schools in 2005 and 2006. He is the editor of *Australian Academic & Research Libraries*.

How long is forever? Paradigms for digital preservation

Abstract

This paper examines and describes some Australian responses to digital preservation, a major issue which currently has few strategies to address it. It presents results of research into two aspects of digital preservation: (1) the changes in preservation paradigms that are required to effectively maintain ongoing access to digital information resources; and (2) some practical implications of the definitions of 'long term' as applied to digital preservation. This paper is based on interviews that were conducted in 2004 and 2005 with senior Australian information management professionals about current digital preservation strategies and issues. The interviewees, who were from library, recordkeeping, audiovisual archiving and geospatial information communities, were in charge of preservation activities at their institutions or were active commentators on digital preservation. This research is based on the contention that Australian practice in digital preservation is at the forefront of world best practice. The findings provide a clearer picture of how preservation is changing in response to the increasingly digital environment. It indicates that a better definition of the timescale for digital preservation assists in determining the resourcing needed to more effectively carry out digital preservation, and suggests ways in which current thinking about digital preservation may need to change.

Introduction

The information community agrees that digital preservation – maintaining access over time to information in digital form – is a significant, but difficult challenge that demands urgent action. As networked computing increasingly becomes part of everyday life, the widespread reliance on digital data is revolutionising how many of us work and play. Librarians and recordkeeping professionals are not exempt from these effects and our traditional practices are changing. Any guidance that allows us to more rapidly develop widely applicable solutions to the challenges of digital preservation is worth putting effort into. Despite notable progress in developing digital preservation strategies, processes and tools, we are still not close to being able to state definitively that we can maintain digital information in a usable form into the future.

Preservation paradigm shifts

The increasingly digital environment that librarians and recordkeeping professionals inhabit is changing work practices and, arguably, the principles behind these practices. The paradigm shift in how libraries and archives carry out their business can be simplistically described as the change from acquiring, storing and providing access to material *in physical forms*, to acquiring, storing and providing access to *digital* material. Preservation activities are significantly affected by the changing paradigm. A paradigm shift is undoubtedly occurring: a senior Australian information professional interviewed as part of this research suggested that ‘the paradigms and the structures are challenging because you’re going from something

you can physically see, touch and deal with to something that is embedded in information management and technology, which deals with different languages etc.’

Our adherence to traditional ways of thinking about preservation is impeding how we address digital preservation. Traditional preservation thinking is based around the conservation of artefacts – physical objects that carry information content – readily apparent from the definitions in the IFLA *Principles for the Care and Handling of Library Materials* (1998). These define the aims of *conservation* as being to ‘slow deterioration and prolong the life of an object’ and the definition of *archival quality* uses the words ‘a material, product, or process is durable and/or chemically stable, that it has a long life, and can therefore be used for preservation purposes’ (Adcock, 1998, p. 4). The traditional preservation paradigm considers information content and the carrier as one and the same, although it must be noted that this understanding was being altered in large-scale copying programs, which often distinguished between content and carrier.

A summary of the basic principles on which the traditional preservation paradigm is based might look like this (not a complete list):

1. The originals (*artefacts*) are preserved so that their information content can be accessed
2. They are preserved for as long a period as possible
3. Preventive conservation (providing appropriate storage and ensuring appropriate handling) is the key to

ensuring the longevity and continuing usability of the artefacts

4. Benign neglect may be an appropriate treatment (based on Cloonan, 1993, p. 596; Harvey, 1993, pp. 14, 140).

Unfortunately, this traditional preservation paradigm cannot be applied in the digital environment, and information professionals have not yet adjusted fully to the new ways of thinking and practice required by a new paradigm. Deanna Marcum, then President of the Council on Library and Information Resources, notes (in Kenney & Stam, 2002, p. v) that

Libraries are working to integrate access to print materials with access to digital materials. There is likewise a challenge to integrate the preservation of analog and digital materials. Preservation specialists have been trained to work with print-based materials, and they are justifiably concerned about the increased complexity of the new preservation agenda.

The traditional preservation paradigm does not offer any understanding of the complexity of copying digital materials, the basis of digital preservation processes. Not just the bit-stream, but also a wide range of other attributes of the digital object needs to be preserved. Additional challenges are presented by the quantities of digital information being produced. New preservation practices and new ways of thinking about preservation are needed.

Two differences between old and new preservation paradigms are of crucial importance. The first is the concept of benign neglect. Paper, photographs and other artefacts are unlikely to be harmed if they are left for some time without intervention; but digital information must be actively maintained from the moment of its creation. Interruptions in its management will quickly result in their becoming inaccessible. Constant maintenance and elaborate 'life-support systems' are required. The second is that for preserving digital material the social and institutional challenges assume much greater importance: 'even the most ideal technological solutions will require management and support from institutions that go through changes in direction, purpose, and funding' (Workshop on Research Challenges in Digital Archiving and Long-term Preservation, 2003, p. 7).

Where Do We Look for Guidance?

Australian digital preservation expertise is widely recognised as at the forefront of world best practice (an interviewee described this as 'Australia generally in this area is ... fighting above its weight') and this is readily demonstrated by reference to the digital preservation literature (for example Beagrie, 2003, p. 14) and the extent to which Australian digital preservation expertise is actively sought.

The experience of the National Library of Australia (NLA) with digital preservation – just one of several Australian institutions who are recognised as at the forefront – is a good example. The NLA has been active since 1994 in digital preservation and shares its expertise and experience willingly with the Australian and international information communities.

In 1995 it founded one of the first library digital preservation sections, set up its archive of online publications, PANDORA, in 1996, and in 1997 established the PADI (*Preserving Access to Digital Information*) Web site. More recently the UNESCO *Guidelines for the Preservation of Digital Heritage* (UNESCO, 2003) were authored by the NLA's Digital Preservation Manager, Colin Webb.

Given this high level of Australian digital preservation expertise, the views of senior Australian information management professionals about current digital preservation strategies were sought to see if their experience and reflections could assist in developing workable strategies for digital preservation. In 2004 and 2005 interviews were conducted with people from the Australian library, recordkeeping, audiovisual archiving, and geospatial information communities who were in charge of the preservation activities of their institutions and/or were active commentators on digital preservation. Institutions represented were the National Screen and Sound Archive, the state libraries of Victoria and Tasmania, VERS (the Victorian Electronic Records Strategy), the Australian Science and Technology Heritage Centre, GeoScience Australia, the Australian Centre for the Moving Image, and the NLA.

Participant' views were sought about definitions, preservation paradigm shifts, digitisation, selection for preservation, strategies, loss or compromise of digital information, threats and challenges, and research and training needs. This paper examines the response of eleven interviewees to questions about how

long we are attempting to preserve digital information for, and about changes in preservation thinking.

Definitions of 'Forever': For How Long Are We Preserving Digital Information?

One question posed during the interviews was 'How long is 'long-term' in digital preservation thinking?' The answers to this question could assist in determining the resourcing needed to effectively carry out digital preservation, and also suggest some ways in which thinking about digital preservation and its processes will need to change. The extremes of the time spectrum have been described by Clifford A. Lynch of the Coalition for Networked Information as, at one end, 'deep time preservation (10,000 plus) and at the other, the next two weeks' (Lynch, 2004).

The shortest period suggested during the interviews was 4-5 years. This is 'long enough for media failure or changes in technology to become an issue' even though 'the vision is for a defined percentage of that digital information to be around for centuries'. Interviewee eleven suggested that our state of knowledge was at present insufficient to achieve longer time-spans so 'we have to think in smaller time-frames in the meantime'. Interviewee 10, from the geosciences community, which closely links its preservation activities to business requirements, suggested 30 to 50 years:

So in terms of long-term for us for business requirements, with the information we're acquiring at the moment, we're possibly thinking up to, say, 30 to 50 years. The things I think of are, say, a

satellite imagery archive and that dates back to 1979, in digital form.

These responses (4-5 years and 30-50 years) were the most conservative supplied by the eleven respondents. The most common response (interviewees 7, 8 and 9) was 100 years as a minimum requirement, although this was justified in different ways. Interviewee 7 suggested economic reasons: 'the return on the investment in digital preservation that we make for me has got to be of a 100-year plus time-span'; this respondent also suggested political reasons – the 100 year period was 'sufficiently long to get over the immediacy of some of the politics surrounding [digital preservation] and firmly establish it'. He suggested that there was another dimension to the question and provided a second answer: 'as long as the community thinks it is valuable to keep'. Interviewee 8 explained 100 years in more pragmatic terms: 'it's the photo of you as the baby that you can look at as the old person'. He considered that this was 'at least a life-span and effectively that's got to approach 100 years ... a life-span where we can still access things of our childhood in old age.' Interviewee 9 suggested that 100 years was notional and that the key was active management of the data:

It's a bit of a moot question, because, if you work on the basis of providing accessibility to content over time then if it's still of value then it's got to be actively managed, so it doesn't really matter if it's a million years.

These were responses from *individuals* who were responsible for setting the

digital preservation agenda for the institutions which employed them. One *organisation's* response, articulated by Interviewee 5, was that they were preserving digital information for 300 years. Setting a specific period informed their current practice:

There's no point in saying forever, because we know that nothing lasts forever, but what we also know is that, in relation to a lot of conservation science you can say, if I do these, then that will be the consequence ... I can say to get to 300 years ... I have to do this, I have to do this other thing ... saying 200 or 300 years informs your current practice. It actually says, well, for it to last 200 or 300 years, I need to do this now, and I need to take these actions now, or I need to find out what would be the appropriate action for me to take for it to be available in 200 years.

Some responses were at the other end of the time spectrum. Interviewee 1 indicated that the only possible answer was 'indefinitely'. Interviewee 3 specified 250,000 years, based on his work with the international nuclear industry and its requirement to be able to access records – 'digital these days'— relating to radioactive waste with its very long half-life. This industry was keen to 'hand to [future generations] everything that they need to know about the past ... so they can then make decisions about what they keep and what they can ... let go.' Interviewee 4 specified that 'I don't have an end-date in my thinking'. His view was formed by his position as a senior manager working in a depository library, where the responsibility for preservation

was commonly considered to be for an indefinite period.

One respondent (Interviewee 1) noted the technical ability to copy digital information without loss as significant in determining how long. Unlike 'analog copying processes which would eventually result in the decay of the information', for digital data

there are no theoretical limitations any more on how long you can keep things going, except things like the death of the sun ... once we get to a point where we have successfully migrated the data through a few generations, we can start to feel fairly comfortable that the data will survive as long as our current society survives.

However, for Interviewee 2 the principal determinant was not the technical preservation processes, but

the organisational context in which these things survive ... you can do all the things that are necessary to continue the physical life of something, but it sits within some sort of organisational context, which we know can be disturbed or changed or destroyed, and, if that context has gone, what happens to the material ... it's a question you can't answer.

Definitions of 'long term' from other sources show similar lack of a clear consensus. In a 1997 report it was defined as 50 years 'as a working hypothesis, a period of time which takes us towards the limits of one of today's most durable storage media: compact disk storage technology' (Bennett, 1997,

p. 5). Bennett comments that 'technology uses an implicitly different timeframe to the accepted principles of preservation, the "technological long term" has a very close or near horizon' (Bennett, 1997, p. 37). 'Long term' is defined in the increasingly influential OAIS (Open Archive Information System) Reference Model (ISO 14721:2003) as 'long enough to be concerned with the impact of changing technologies, including support for new media and data formats, or with a changing user community. Long Term may extend indefinitely' (Consultative Committee for Space Data Systems, 2002, p. 1-1). This definition accords with some of the definitions provided by interviewees, not surprising as the OAIS model is gaining currency as a framework for many digital preservation activities.

Is There a Paradigm Shift?

The interviews posed three questions intended to solicit views about whether there was a paradigm shift in preservation and about its characteristics:

- How has preservation changed over time in your profession? In your experience? In your institution?
- In your profession, has the preservation mind-set changed sufficiently to accommodate the preservation of digital information?
- To what extent are 'old preservation paradigm' (i.e. artefact-focussed) ways of thinking still paramount?

Most respondents agreed that there had been a change of paradigm or that we were in the middle of a paradigm change, but some did not. Interviewee 7, who identified himself as 'a library preservation administrator', suggested

that from his perspective 'I don't think ... the fundamental concepts have changed'. He did acknowledge that there had been a change from emphasis on 'single-item salvation' to a collection-wide perspective. Interviewee 8 bluntly and pessimistically suggested that 'there isn't a preservation mindset'. In his opinion, 'we're just getting used to this loss.'

The respondents did not, however, agree about the nature of the changes. There was general agreement that the old paradigm was based on the conservation and preservation of artefacts, Interviewee 1 expressing this as 'attempting to do perfect preservation' in the past and Interviewee 6 describing it as about 'repairing damaged physical records basically'. Several respondents indicated that there was still a place for artefact-based thinking in the increasingly digital environment. Interviewee 5 suggested that traditional paradigm skills and ways of thinking would still be necessary in the digital environment 'where we're coping with legacy data ... we keep finding old stuff, old electronic stuff in our holdings, so there is an element of having to do that, but it's almost become a sort of antiquarian interest.'

Interviewee 5 noted the difficulty in changing habitual practices. Many working in the information professions found it hard to alter their ways of thinking and working: there was, he suggested, 'a degree of difficulty for practitioners in a physical space in being able to bridge the gap to dealing with digital content'. This was not because 'they don't want to'; rather, it was because 'the paradigms and the

structures are challenging because you're going from something you can physically see, touch and deal with to something that is embedded in information management and technology, which deals with different languages etc.' Interviewee 3, from the archives community, suggested that traditional thinking was still being applied in the digital environment but was not helpful:

we see it in archivists and a lot of members of the profession in Australia and archivists overseas ... that they still keep falling back on the life of a CD, the life of a floppy disk, the life of an optical medium ... and they haven't really thought about all or addressed the issues of what it really means to be digital and also that they are living in an open network environment and what this can mean for the way they can preserve and manage things.

This is changing, according to Interviewee 5, who suggested that by the late 1990s the archives community recognised that the 'problem is a data problem, it's not a software application problem'.

Only one respondent (Interviewee 1) asserted that the old paradigm thinking had little effect:

most people who have even the most preliminary steps in managing digital information would realize that preserving the physical carrier is not what you're trying to do ... Preserving digital information requires an entirely different approach to traditional preservation, not just different skills.

However, old paradigm thinking should not be totally discarded because ‘there are relevant analogies at the highest level (e.g. assigning priorities, undertaking appraisal and retention, reformatting to more stable media etc.)’ as well as major differences.

What Constitutes the New Preservation Paradigm?

What, then, did respondents consider the elements of the ‘new’ preservation paradigm to be? There was agreement on three points:

- an increased emphasis on risk management and on intellectual control, and a dependence on high-tech infrastructure;
- the relation between preservation and access will be more closely linked; and
- we will develop new languages to describe what we do and reach new understandings of data and its relationships.

Interviewee 10 saw no place for the original artefact in the new digital paradigm. Access was the reason:

Our whole business is about generating and delivering information ... that’s our vision statement ... so that’s why the actual access to it is really important to us. If we’ve got all this fantastic information that we’ve gathered that is sitting in an inaccessible form, it’s of very limited value ... you know, we’re not fulfilling our mission.

This was a minority view, several others commenting (although not

unequivocally) on the value of artefacts. Interviewee 1 noted that in audio-visual archives ‘there are still conflicts and tensions about things like the need to always retain the original artefact which is a fairly firm principle’. This could change: ‘we may dispose of original materials once we are adequately satisfied that the digitisation has captured all of the properties that we want to capture’.

Interviewee 2 considered that there had been a change from the situation where ‘thirty years ago we were thinking mainly of the content without knowing that’s what we were doing’ to one where ‘we are much more aware now of the artefact value of certainly moving images and sound recordings, and the difference between the carrier and the content’. This was redefining preservation practice from a primary emphasis on copying to an awareness that the original carrier needs to be kept ‘for its life, whatever that is, so that we don’t close the door on other possibilities’. Two respondents suggested that preservation paradigms would change significantly but we would keep returning to old paradigm practices for solutions to digital preservation challenges. There was still potential in the development of extremely durable carriers: silicon carbide (Interviewee 3) and

a really good area of research is actually to go ... look at how we can build objects, whether they’re crystalline structures or biosomethings or a kind of digital synrock if you like ... whether we can actually lock these, cement these 0s and 1s into some

substance, something that is stable through time (Interviewee 8).

One strong conclusion to be drawn is that new preservation paradigms were largely about ways of thinking rather than about technical solutions. Only one respondent (Interviewee 2) specifically noted a technical aspect, the increased 'reliance on an industrial infrastructure which becomes ever more complex'. He expressed concern about 'what do you do if that collapses'. The rest of the responses were about changes in understanding, in how to describe what we deal with, and in placing different emphases on aspects of current practice.

Several respondents mentioned risk management and intellectual control. Although these are already present in traditional preservation thinking, they would need to receive considerably greater emphasis. Interviewee 7 suggested that risk management was in fact 'the business of preservation ... it's about taking ... a risk approach, which is what is at greatest risk in that particular collection and what can we do to control the risk to try to get a better outcome for as long as we possibly can'. Interviewee 1 was less emphatic: the new preservation paradigm has 'an increased emphasis on a risk management approach ... that means both identifying potential risks more comprehensively and taking action within a whole range of different ways to manage those risks. It also does mean accepting some risk as inevitable.'

Two respondents considered that improved intellectual control was the key to effective digital preservation. One (Interviewee 3) suggested that

the only way to manage [digital preservation] is to have an intellectual control system that applies across all media ... the intellectual control system is a sort of conceptual system above whatever technology you use.

Changing Ways of Thinking

Elements that will make up the 'new' preservation paradigm were suggested by the respondents, who focused on the relation between preservation and access, the need for a new language to describe what we do, and developing new understandings of data and its relationships. For one respondent (Interviewee 6) preservation had become the dominant way of thinking about ensuring access. His archive had recently 'given preservation much more prominence in the organization'. Here 'the digital preservation work that we've been doing is funded out of that funding as well, because it is aimed at preserving digital records and making them accessible'.

Interviewee 9, also from an archives environment, suggested that digital preservation practices needed 'a different language and concept' where we need 'language translation. It's the ... big ball on the beach which has two sides, black and white. On one side we've got one person saying it's black and the other says white, but if you're up on the cliff you can see both sides.' We need to be more flexible: 'do what we can practically do now, but watch what's emerging and be in a state that we can incorporate or move into that different way of doing things without having to actually go back and restart'.

A new understanding of how software interacts with data was emerging. Interviewee 5, from the archives context, described this gradual realisation:

We then ... came to realise, well, at the end of the day records are just data and data are just records and in that process we came more to a clearer understanding of how software applications interact with data to produce records ... that sort of radically changed the possibilities of finding solutions and we found the solution very quickly ... basically we really started to give serious attention to this in the year 2000 and we had our basic approach worked out within a year. The basic idea of normalising materials into XML structures and ensuring that those XML structures have appropriate metadata kept with them, so that ... the provenance and all the other authenticity issues about them could be re-established.

How will we recognise when we have a new preservation paradigm? Interviewee 4 suggested when digital practice is fully integrated into standard operational practice. Until this is resolved, digital preservation will continue to be a problem:

the paradigm hasn't shifted to the extent where digital objects and content is incorporated completely within the normal processes that the library undertakes. So they are treated as a different acquisition process, or as a different this or a

different that and, I think, until we get that integration, we're going to have this sort of difficulty with digital preservation ... it's a competitive world within libraries, but still traditional services seem to speak very loud at the budget table.

Lessons for Professional Practice

The respondents provide us with a better, although still incomplete, picture of how preservation is changing in response to the digital environment. Their responses are based on actual experience, thus making their observations valuable. Such experience is rare: a report of the digital preservation activities of institutions from 13 countries in 2004 noted that of the 48 completed surveys 'less than a quarter of respondents had actual production experience in implementing active preservation strategies' (OCLC/RLG PREMIS Working Group, 2004, p. 53).

Indications provide by the respondents assist by better defining the timescale that can help us to more precisely determine the resourcing needed for effective digital preservation and by suggesting some ways in which our thinking and procedures need to change. For example, for one respondent setting a specific period (300 years) informed the current practice of their institution. For another 100 years as a minimum requirement was decided on for pragmatic economic reasons, this period being considered long enough to allow a return on investment.

The responses also allow some of the implications of changing ways of

thinking about preservation to be more clearly described. The traditional preservation paradigm will still play an important role but will no longer be predominant. Preservation of digital materials places considerably less emphasis on technical aspects, and significantly more on characteristics such as risk management principles and intellectual control through the use of metadata. Clearer definitions are required, examples being of the relationship between preservation and

access, and how software interacts with data. There will still be emphasis in some circumstances on the importance of keeping original carriers, and there is some potential for developing very durable, preferably machine-independent carriers

Although it is premature to make definite statements about the implications of these findings for professional practice, they do allow us to begin to see what changes are in store.

Acknowledgements

I owe thanks to the interviewees who gave freely of their time and expertise, and to John Mills and Stuart Ferguson, Charles Sturt University, for commenting on an earlier version of this paper. Earlier versions of part of this paper were presented at the 2004 LIANZA conference (Harvey 2004) and appeared in *Preserving Digital Materials* (Harvey 2005).

References

- Adcock, E.P. (1998). *Principles for the care and handling of library materials*. Paris: IFLA PAC.
- Beagrie, N. (2003). *National digital preservation initiatives: An overview of developments in Australia, France, the Netherlands, and the United Kingdom and of related international activity*. Washington, D.C.: CLIR and Library of Congress.
- Bennett, J.C. (1997). *A framework of data types and formats, and issues affecting the long term preservation of digital material*. London: British Library Research and Innovation Centre.
- Cloonan, M.V. (1993). The preservation of knowledge. *Library Trends* 41(4), 594-605.
- Consultative Committee for Space Data Systems. (2002). *Reference model for an open archival information system (OAIS)*. CCSDS 650.0-B-1 Blue Book. Washington D.C.: NASA.
- Harvey, R. (2004, September 8). The preservation of documentary digital heritage: Lessons from Australian experience. Paper presented at the *Made in Aotearoa LIANZA Conference 2004*.
- Harvey, R. (1993). *Preservation in libraries: Principles, strategies and practices for librarians*. London: Bowker-Saur.
- Harvey, R. (2005). *Preserving digital materials*. München: K.G. Saur.

- Kenney, A.R. & Stam, D.C. (2002). *The state of preservation programs in American college and research libraries: Building a common understanding and action agenda*. Washington, D.C.: Council on Library and Information Resources.
- Lynch, C. (2004). Editor's interview with Clifford A. Lynch. *RLG DigiNews*, 8(4), Retrieved July 10, 2006 from http://www.rlg.org/en/page.php?Page_ID=19481
- OCLC/RLG PREMIS Working Group. (2004). *Implementing preservation strategies for digital materials: Current practice and emerging trends in the cultural heritage community*. Dublin, OH: OCLC.
- UNESCO. (2003). *Guidelines for the preservation of digital heritage*. Prepared by the National Library of Australia. Retrieved July 10, 2006, from <http://unesdoc.unesco.org/images/0013/001300/130071e.pdf>
- Workshop on Research Challenges in Digital Archiving and Long-term Preservation. (2003). *It's about time: Research challenges in digital archiving and long-term preservation: Final report*, Workshop on Research Challenges in Digital Archiving and Long-term Preservation. Washington, D.C.: National Science Foundation and Library of Congress. 51.