A BRIEF INTRODUCTION TO DIGITAL PRESERVATION

Steve Knight*
Digital Library Transition Co-ordinator
National Library of New Zealand
P O Box 1467
Wellington
New Zealand
Ph: 04 474 3142
Fax: 04 474 3140
Email: steve.knight@natlib.govt.nz

Introduction

In the interests of brevity it is reasonable to put a stake in the ground and claim that serious debate around '*the death of the digit*'[1] began with publication of the Commission on Preservation and Access (CPA) and Research Libraries Group's paper *Preserving Digital Information*[2] in 1996.

However, a year earlier the Joint Information Systems Committee (JISC) and the British Library had organised a conference on *Long term preservation of electronic materials*[3] at Warwick which was chaired by the now Chief Executive of the British Library Lynne Brindley. The outcome of the Warwick conference (in the light of the soon to be released CPA/RLG document) was the commissioning of a number of studies which have provided a benchmark for subsequent work on digital preservation.[4]

The major findings of the CPA/RLG report (all of which are still valid and being worked on today) included the following:
- The first line of defense against loss of valuable digital information rests with the creators, providers and owners of digital information.
- Long-term preservation of digital information on a scale adequate for the demands of future research and scholarship will require a deep infrastructure capable of supporting a distributed system of digital archives.
- A critical component of the digital archiving infrastructure is the existence of a sufficient number of trusted organizations capable of storing, migrating and providing access to digital collections.
- A process of certification for digital archives is needed to create an overall climate of trust about the prospects of preserving digital information.[5]

Digital preservation issues

Digital material requires new methods of storage, management and presentation. Not surprisingly digital preservation issues fall into a number of familiar categories – technical, organisational, legal and, maybe less obviously, social.

Technical issues

Digital material is fragile, ephemeral and poorly suited to being retained 'in perpetuity' to use the language of the National Library's new Act. Technological change manifests itself in a number of ways, eg. the change from 5 ¼ to 3 ½ inch floppy disks, the seven iterations of MS Word from Word 2 (1991) to Word 2003 (with limited backwards compatibility) and the countless software programmes that are just no longer available such as the early word processing package WordStar initially released in 1979.

Hardware, software and other devices (eg. printers) have an average life cycle of between 2 and 5 years. Despite valid concern about the permanence and stability of physical digital media it is highly likely that these media will in fact outlast the hardware, software and other devices needed to use them.

Preservation of physical media then becomes a lesser concern than the general problem of technological obsolescence.

Technological obsolescence has generated a range of responses (the following definitions have been taken from Maggie Jones and Neil Beagrie's handbook of preservation management):[6]
- Technology preservation – preserving the original software (and possibly hardware) that was used to create and access the information
- Refreshment – copying information content from one storage media to the same storage media
- Reformatting – copying information content from one storage medium to a different storage medium (media reformatting) or converting from one file format to a different file format (file reformatting)
- Migration – transferring digital resources from one hardware/software generation to the next
- Emulation – developing techniques for imitating obsolete systems on future generations of computers.

Each of these has advantages and disadvantages and it is not yet clear what the final approach will entail although a mixture of migration and emulation is beginning to look like a feasible strategy.

Organisational issues

'In addition to redefining responsibilities of organisations, it may be necessary to redefine roles within organisations to ensure long-term access to digital information … for example, responsibility for maintaining long-term access to digital records may be shared between business managers, records management and information technology personnel, and individual creators.'[7]

Important questions include:
- How do we make decisions to allocate resources to digital preservation when other aspects of our response to the digital world – digitisation, site licensing etc – are competing for the same funds in what is often a static funding environment?
- What are the legal implications of the various strategies for digital preservation, almost all of which require copying of some form or another?

- Where do we find the staff that we need to implement digital preservation strategies? With technological change proceeding as fast as it is who is able to train people with the appropriate skills?

It is imperative that digital preservation is embedded in our daily activities and that we are all aware of it as an integral component of our work as information professionals. Although almost ten years old the following quotes are still apposite today. In 1995 Peter Graham claimed that '*nothing makes clearer that a library is an organisation, rather than a building or a collection, than the requirement for institutional commitment if electronic information is to have more than a fleeting existence*'[8]. Paul Conway reinforced this the next year when he noted that 't*he real challenge is creating appropriate organisational contexts for action*'.[9]

Legal issues

Important questions include:
- What is the relationship of copyright legislation to licensing (including shrink-wrapped licenses) and contract law, eg. some rights available under copyright are lost under licensing.
- As licensing does not entail ownership how do we ensure contractual agreements to maintain access to material when the license period ends and whose responsibility is the preservation of that material?
- What is the legal status of the preservation strategies for technological obsolescence outlined above? What is the impact, not only on content, but on the hardware on which the content resides and the software required to read the content?
- How will legal deposit of electronic materials be effected? What are the mechanisms that will be needed to implement legal deposit of electronic materials? How are relevant roles and responsibilities to be delineated?
- What are the compliance implications of negotiated usage restrictions and statutory requirements such as the Privacy Act or the Official Information Act for digital repositories?

These are not just questions for the organisation responsible for digital preservation. Over time the answers must include the creators of the information, the owners of the intellectual property in the resources (if that is different from the creator) and the potential users of the information.

Of more fundamental concern may be the need for the preserving organisation to be seen as acting in a manner that will instil faith in their ability to perform the task that has been set for them.

When we take into account the borderless nature of digital information the potential pitfalls become even clearer.

Social issues

Three key issues to be canvassed here are capability, cost and redundancy.

Capability: There is an increasing acceptance that digital preservation may not be the province of a single organisation. There are probably few organisations, even

internationally, who have the mandate, let alone the technical, staffing and financial resources to be able to develop a sustainable 'trusted digital repository'.

Cost: It is difficult to say with any certainty what the costs of digital preservation are going to be.

A recent report from the National Library of Australia notes that '*a surprising observation … was that with one or two exceptions, national libraries have done very little long-term corporate planning for their new roles in the digital age. Most recognise that they have inadequate technical infrastructure in place to support their digital collections but are unsure what to do about this. There was little evidence of attempting to integrate new activities and roles into strategic planning or mainstream operations, and there is no understanding of the costs entailed in digital archiving.*'[10]

Similarly, a recent review of National Library of New Zealand digital archiving activities found that '*despite numerous attempts to quantify the costs of building digital libraries the costs of selection, acquisition, ingest, and cataloguing of digital content remain a matter of guesswork. Where organisations have attempted to produce detailed costings they have done so mainly at the macro level and against an array of assumptions and guesses that can not easily be verified or replicated.*'[11]

Redundancy: The objects we collect will increasingly be created on computers, collected from computers, stored in computers, preserved in computers and made accessible from computers. As a result the need for redundancy will increase.

The review of the National Library of New Zealand's digital activities quoted above also noted that '*as the digital holdings of the Library continue to expand and begin in their number and extent to reflect the prevalence of digital documents in society, their loss would have an increasingly catastrophic impact on the Library's core activities as well as on the record of the cultural and scientific heritage of New Zealand and the South Pacific … The Library should ensure that there is a level of distributed redundancy in its systems to ensure that the loss of one location would not put its entire digital library at risk.*'[12]

What's going on at the moment?

The following is not meant to be exhaustive. It is designed to give a general sense of what is going on in the digital preservation space.

Consciousness raising: Awareness of the potential for loss of digital materials is becoming more visible on the world stage. The European Commission[13] has provided funding for dissemination of information on digital preservation through ERPANET[14] while UNESCO is presenting a draft charter[15] for consideration at its 32nd General Conference in Paris this year and has sponsored the writing of a set of guidelines for digital preservation by the National Library of Australia.[16]

Preservation metadata: Development of conceptual and practical preservation metadata schemas has been ongoing during the last five or six years. This includes work by the National Library of New Zealand.[17] The NLNZ preservation metadata schema was developed from the experience of other initiatives, in particular work undertaken by the National Library of Australia[18], the CEDARS programme[19], OCLC/RLG activities[20] and

the emerging consensus regarding the role of the OAIS Reference Model.[21] Further work is being undertaken at the National Library of New Zealand in developing an implementation ready data model[22] along with an XML schema version of the data model.[23] This work complements that of the PREMIS Working Group organised by the Research Libraries Group to look into preservation metadata implementation strategies.[24]

Certification: The notion of certification of digital repositories has gained traction over the last 12-18 months with OCLC and RLG again at the forefront of research into what will be required for an organisation to be able assert 'trusted digital repository' status for itself.[25]

File formats: Concern over the file format types that can be used for digital preservation and their ongoing viability, particularly through multiple iterations, was flagged as early as 1999[26] but is gaining in visibility[27] with the Digital Library Federation looking into the possibility of a Global Digital Format Registry[28] and the Public Records Office's PRONOM (a database designed to store information on software applications and their related file formats).[29]

Applications: The first digital repository applications are starting to become available for testing by organisations engaged in digital preservation. They include DSpace[30] and FEDORA (Flexible Extensible Digital Object and Repository Architecture).[31] JISC, in the UK, are currently inviting bids to '*to establish a national Digital Curation Centre (DCC) to lead research and development into key areas of digital curation for data and publications, and to pilot the development of generic support services for maintaining digital data and research results over their entire life-cycle for current and future users.*'[32]

The National Library of New Zealand response

Conservative estimates suggest that by 2005 there could be 12 Terabytes of unique, digital original material available online in New Zealand, with a growth rate upwards of 1 Terabyte annually.[33] Paralleling this, a recent survey in New Zealand found that while storage costs are falling approximately 35% per year any savings from cheaper storage are being immediately overtaken by a growth in storage requirements of up to 50%.[34] To put this into perspective Papers Past is already nearing 700GB or almost 1200 CD-ROMs.

The National Library of New Zealand has been developing its approach to the long-term management of electronic material and development of a Digital Archive over the last 2-3 years. The use of the word 'Archive' here is deliberate and follows the usage of the Task Force on Archiving of Digital Information mentioned above in which '*digital archives are distinct from digital libraries in the sense that digital libraries are repositories that collect and provide access to digital information, but may or may not provide for the long-term storage and access of that information*'[35].

This work has received further impetus with the passing of the National Library of New Zealand (Te Puna Mātauranga o Aotearoa) Act 2003, which for the first time gives the Library a mandate to collect, preserve, protect and make accessible electronic resources, both online and offline. Part 29(1) defines an electronic document as '*a public document in which information is stored or displayed by means of an electronic recording device, computer or other electronic medium, and includes an internet document*' and an internet document as '*a public document that is published on the internet, whether or not*

there is any restriction on access to the document; and includes the whole or part of a website'. Note that a public document is defined elsewhere in the Act.

It is within this framework that NLNZ is undertaking a programme of linked initiatives to ensure the incorporation of digital material into the Library's core business processes with a view to the long-term accessibility of those resources. The goal of the programme is to develop holistic end-to-end processes for the handling of digital material.

The programme includes the following activities:
- Developing and implementing business process workflows for incorporating digital objects into the Library's business processes, eg. selection, acquisition, care and handling, transformation of physical originals
- Developing infrastructure for digital material, eg. storage, authentication, access
- Purchasing and implementing a metadata repository for provision of portal services to NLNZ applications
- Researching and implementing digital library 'components', eg. preservation metadata (schema, extraction, storage), persistent identifiers
- Piloting web harvesting for the capture and preservation of New Zealand based and related web sites.

The progress of the Library to date has thrown up a number of major areas of need that will require continued attention if we are to successfully confront the challenge of digital preservation. These include:
- Recognition that while information in all formats is still increasing more and more is being produced digitally and the gap is constantly increasing
- Engagement with the wider information community will become increasingly important as it is unlikely that one organisation is going to be able to do it all
- Allocation/reallocation of resources to digital preservation and developing the appropriate skill base
- Ensuring we have the necessary technology infrastructure
- Development of appropriate strategies, policies, processes and procedures
- Ensuring our selection, acquisition, and description processes are in sync with the requirements of digital preservation.

Trusted Digital Repositories

Implicit in the figures and initiatives described above is the need to provide a 'trusted repository' for digital material which ensures its viability and authenticity over time.

The 1996 Task Force Commission on Preservation and Access paper mentioned earlier recognised that '*for assuring the longevity of information, perhaps the most important role in the operation of a digital archives is managing the identity, integrity and quality of the archives itself as a trusted source of the cultural record. Users of archived information in electronic form and of archival services relating to that information need to have assurance that a digital archives is what it says it is and that the information stored there is safe for the long term*'.[36]

The notion of a trusted digital repository has been most clearly articulated by RLG and OCLC[37]. The RLG/OCLC attributes of a trusted digital repository comprise:

- Compliance with the Reference Model for an Open Archival Information System (OAIS) administrative responsibility
- Organizational viability
- Financial sustainability
- Technological and procedural suitability
- System security
- Procedural accountability.

It is interesting to note that four of these attributes (responsibility, viability, sustainability and accountability) are more social than technical measures.

While trust is already a feature of the Library in its capacity as a national library, it is not a given that trust will automatically be bestowed upon it in the digital arena. The Library's work with digital material needs to leverage off its status of trust in the analogue context but it must develop a reputation for trustworthiness in the digital arena over time through transparency of process, accountability and reliability.

Given New Zealand's small population base and the few agencies of appropriate size, funding and mandate that could possibly take on the role of a 'trusted repository' it is likely that the Library will need to look at undertaking this role in conjunction with other agencies. Therein lies a huge challenge both for the Library and for the wider information community.

Conclusion

This paper has briefly looked at:
- key early work on digital preservation
- the range of issues to be addressed
- current activities underway internationally
- National Library of New Zealand's efforts, and
- the wider world of digital repositories and the notion of the trusted repository.

The desired outcome of all of these activities, local and international, is the development of programmes to ensure the long-term preservation of digital materials.

There is no doubt that we will face many challenges in ensuring that digital objects remain functional into the future. One aspect of meeting this challenge may be a move away from the defeatist language that describes the requirements of digital preservation as 'problematic' and the accumulation of digital material as 'an unmanageable flood'.[38]

Digital preservation is not something we can shy away from even if it is far from clear what is required now nor what will be required as we move to ensure that digital preservation is an integral component of our day-to-day activities in the future.

At the risk of being repetitive I'm going to finish with another quote from the 1996 Task Force which rightly stated that '*the problem of preserving digital information for the future is not only, or even primarily, a problem of fine tuning a narrow set of technical variables. It is not a clearly defined problem … rather, it is a grander problem of organizing ourselves over time and as a society to manoeuvre effectively in a digital landscape. It is a*

problem of building … the various systematic supports … that will enable us to tame anxieties and move our cultural records naturally and confidently into the future.'[39]

References

[1] Feeney, Mary. (1999). Digital Culture: maximising the nation's investment: Report of seminar. National Preservation Office. http://www.bl.uk/services/preservation/report.html. Accessed 20 September 2003.

[2] Waters, D. & Garrett, J. (1996). *Preserving digital information. Report of the Task Force on Archiving Digital Information commissioned by the Commission on Preservation and Access and the Research Libraries Group*. Washington, D.C.: Commission on Preservation and Access. http://www.rlg.org/ArchTF/index.html. Accessed 20 September 2003.

[3] Fresko, M. (1996). *Long term preservation of electronic materials. A JISC/British Library workshop as part of the Electronic Libraries Programme (eLib), organised by UKOLN, 27-28 November 1995, University of Warwick. British Library R&D Report 6328*. London: The British Library. http://www.ukoln.ac.uk/services/papers/bl/rdr6238/. Accessed 20 September 2003.

[4] Feeney, 1999, p 74.

[5] Waters & Garrett, 1996, p 40.

[6] Jones, M. & Beagrie, N. (2001). *Preservation management of digital materials: A handbook*. London: British Library.

[7] National Library of Australia. PADI: Preserving access to digital information: Roles and responsibilities. http://www.nla.gov.au/padi/topics/8.html. Accessed 20 September 2003.

[8] Graham, Peter S. (1995). *Requirements for the Digital Research Library*. http://www.ifla.org/documents/libraries/net/drc.htm. Accessed 17 September 2003.

[9] Conway, Paul. (1996). *Preservation in the Digital World*. http://www.clir.org/pubs/abstract/pub62.html. Accessed 17 September 2003.

[10] Gatenby, Pam. (2002). *Report on senior executive fellowship to research digital archiving in national libraries*. http://www.nla.gov.au/nla/staffpaper/2002/elect.html. Accessed 20 September 2003.

[11] Ross, S. (2003). *National Library of New Zealand Te Puna Mātaurange o Aotearoa: Digital library development review*. http://www.natlib.govt.nz/files/ross_report.pdf. Accessed 20 September 2003.

[12] Ibid. p 27.

[13] European Commission. (2002). *Council resolution of 25 June 2002 on preserving tomorrow's memory – preserving digital content for future generations*. http://europa.eu.int/eur-lex/pri/en/oj/dat/2002/c_162/c_16220020706en00040005.pdf. Accessed 21 September 2003.

[14] http://www.erpanet.org. Accessed 21 September 2003.

[15] UNESCO. (2003). *Draft charter on the preservation of the digital heritage. General Conference 32nd session, Paris 2003, 32 C/28*. http://unesdoc.unesco.org/images/0013/001311/131178e.pdf. Accessed 21 September 2003.

[16] Webb, C. (2003). *Guidelines for the preservation of digital heritage*. Paris: UNESCO. http://portal.unesco.org/ci/ev.php?URL_ID=8967&URL_DO=DO_TOPIC&URL_SECTION. Accessed 21 September 2003.

[17] National Library of New Zealand. (2003). *Metadata Standards Framework- Preservation Metadata (Revised)*. http://www.natlib.govt.nz/files/4initiatives_metaschema_revised.pdf. Accessed 16 September 2003.

[18] National Library of Australia. (1999). *Preservation Metadata for Digital Collections – Discussion Draft*. http://www.nla.gov.au/preserve/pmeta.html. Accessed 16 September 2003.

[19] Cedars: CURL exemplars in digital archives. (2002). Cedars guide to: Preservation metadata. http://www.leeds.ac.uk/cedars/guideto/metadata/guidetometadata.pdf. Accessed 16 September 2003.

[20] OCLC/RLG Preservation Metadata Working Group. http://www.oclc.org/research/projects/pmwg/. Accessed 16 September 2003.

[21] Consultative Committee for Space Data Systems. (2001). *Reference model for an Open Archival Information System (OAIS)*. http://www.ccsds.org/documents/650x0b1.pdf. Accessed 16 September 2003.

[22] National Library of New Zealand. (2003). *Metadata Standards Framework- Metadata Implementation Schema*. http://www.natlib.govt.nz/files/nlnz_data_model.pdf. Accessed 16 September 2003.

[23] National Library of New Zealand. (2003). http://www.natlib.govt.nz/en/whatsnew/4initiatives.html#meta. Accessed 16 September 2003.

[24] OCLC/RLG Preservation Metadata Working Group. *Working group II: Implementation strategies*. http://www.oclc.org/research/projects/pmwg/. Accessed 21 September 2003.

[25] Research Libraries Group. *Task force on digital repository certification*. http://www.rlg.org/longterm/certification.html. Accessed 21 September 2003.

[26] Ross, S. & Gow, A. (1999). *Digital Archaeology: Rescuing neglected and damaged data resources*. London: Joint Information Systems Committee and the National Preservation Office. http://www.ukoln.ac.uk/services/elib/papers/supporting/pdf/p2.pdf. Accessed 21 September 2003.

[27] Leeds University. (2003). *Survey and assessment of sources of information on file formats and software documentation: Final report*. http://www.jisc.ac.uk/uploaded_documents/FileFormatsreport.pdf. Accessed 21 September 2003.

[28] Abrams, S.L. & Seaman, D. (2003). *Towards a global digital format registry*. Paper presented at World Library and Information Congress: 69th IFLA General Conference and Council, 1-9 August 2003, Berlin. http://www.ifla.org.sg/IV/ifla69/papers/128e-Abrams_Seaman.pdf. Accessed 21 September 2003.

[29] Public Record Office. *Digital preservation: PRONOM*. http://www.pro.gov.uk/about/preservation/digital/pronom/default.htm. Accessed 22 September 2003.

[30] http://www.dspace.org/. Accessed 21 September 2003.

[31] http://www.fedora.info/. Accessed 21 September 2003.

[32] Joint Information Systems Committee. (2003). *Circular 6/03 (Revised) Digital Curation Centre: Deadline for proposals*. http://www.jisc.ac.uk/index.cfm?name=funding_digcentre. Accessed 22 September 2003.

[33] Extrapolations from: Lyman, P. & Varian, H. (2000). *How much information?* Berkeley, CA: University of California, Berkeley. http://www.sims.berkeley.edu/research/projects/how-much-info/. Accessed 20 September 2003.

[34] Braddell, R. (2002). *Data storage needs escalate.* Wellington, N.Z.: The Dominion Post, Monday October 14, p C10.

[35] Waters & Garrett, 1996, p iii.

[36] Waters & Garrett, 1996, p 23.

[37] Research Libraries Group. (2002). *Trusted digital repositories: Attributes and responsibilities: An RLG-OCLC report*. http://www.rlg.org/longterm/repositories.pdf. Accessed 16 September 2003.

[38] University of Heidelberg Institute for Chinese Studies. *Digital Archive for Chinese Studies: About DACHS*. http://www.sino.uni-heidelberg.de/dachs/intro.htm. Accessed 22 September 2003.

[39] Waters & Garrett, 1996, p 7.