Getting the whole picture: Finding a common language between digital preservation and conservation

Douglas Elford, Lisa Jeong-Reuss, Somaya Langley and Melanie Wilkinson

Abstract

From the outset, digital preservation and conservation seem to be two very different professions. This paper explores the idea that both traditional and digital conservators not only have many commonalties but also an invaluable knowledge bank of their own areas of expertise, which should be shared between these seemingly distinct communities. Preservation Services and the Digital Preservation team from the National Library of Australia (NLA) discuss the commonalities between the two professions. Digital preservation is a relatively young and innovative discipline, while conservation has undergone its own phases of growth and innovation to become a profession that is steeped in tried and tested methodologies and techniques. This analysis of the common challenges shared by the Conservation team and the Digital Preservation team at the NLA illustrates the lessons that can be learned and shared between the two. With multiple linkages between digital preservation and conservation, the common ethos that both communities are committed to preserving culture and heritage emerges and is defined. Practitioners working in conservation and digital preservation need to commit to a holistic approach to preserving collections. Preservation practitioners all work towards ensuring the longevity of cultural collections using a variety of methods, approaches, tools and principles, and only the formats of these collections differ.

Keywords

digital preservation, conservation, significance, digital stewardship, responsible custodianship, authenticity, accessibility, PANDORA, Prometheus, Mediapedia

Introduction

Digital culture, digitisation and digital preservation are all common buzzwords that we've frequently heard. Now, more than ever, it is important to engage more thoroughly in discussions about the preservation of digital materials, introducing necessary terms into the common vocabulary in order to articulate the challenges presented by this rapidly expanding area of cultural collections. Collection custodians are responsible for the care of the collection as a whole, including both physical and digital items. In spite of the intangible and at times ephemeral nature of digital collections, the fundamental purpose driving both digital preservation and conservation are conceptually quite similar.

In traditional conservation significance, authentication, accessibility and responsible stewardship are concepts that are keenly understood by conservators and put into practice on a daily basis for physical collection materials (Keene 2001). Highlighting commonalities between digital preservation and conservation, this paper explores selected approaches to digital content management and the feasibility of these approaches within collecting institutions, with reference to some of the current activities at the National Library of Australia (NLA).

As conservators, it is pertinent that we begin to address the challenges of digital collecting, management and preservation. In recent years there have been studies on copyright law relating to digital preservation as well as changes to legal deposit legislation in various states (such as in Tasmania and the Northern Territory) to incorporate electronic materials. In early 2012 the Australian government proposed a reform 'to extend the legal deposit obligation in section 201 of the Copyright Act 1968 in relation to material deposited with the National Library of Australia' (Attorney-General's Department 2012) to include electronic 'library material'. Progress is also proposed at an international level, where the 2012 review of the UNESCO Charter on the Preservation of Digital Heritage intends to extend its guidelines to include information on the provision of online access to digital collections as well as focusing on achievable preservation strategies for developing nations. These discussions happening internationally illustrate that preservation and the provision of access to born-digital and digitised content and culture are paramount. The digital revolution has long since occurred, yet collection policies that suit digital content in the networked environment are, for the most part, yet to be put into practice. While some new approaches to collecting digital materials in a proactive manner (including more frequently and via semi-automated mechanisms) are required, it is imperative that the field of digital preservation also borrows from long-established collection and conservation processes and practices that have been refined over decades by preservation professionals in Australia.

As digital content often takes on many hybrid forms, the challenges associated with digital collecting and management cross many boundaries. These materials demand expertise often from various different divisions within organisations as well as

between institutions. Acquiring, managing, preserving and providing access to digital culture is a challenge that is faced by all cultural and heritage organisations worldwide – from historical societies to national institutions. Creators, curators, producers, collection managers, cataloguers and conservators all have an ethical responsibility to identify and accept the challenges of twenty-first century collecting. The preservation of cultural heritage in all its perceived intangible and tangible forms is likely to demand collaborative effort. The heritage and collecting community as a whole is responsible for comprehending and accurately articulating the challenges of digital collecting, which will impact us either directly or indirectly.

Significance

Significance is a changeable paradigm for everyone concerned in collections care and management. Those who have contributed to significance statements or registers of significance appreciate the complexity and sensitivity required in undertaking such tasks. Significance is a fluid concept, changing over time and with additional knowledge or context (Russell and Winkworth 2009). It is, however, only one of several drivers behind conservators' approaches to treatment of collection materials. Constant considerations of time and budget allocation for the treatment of each object, as well as the storage methods that can be provided, complement the significance of an item (Russell and Winkworth 2009). Such considerations are equally pertinent for digital objects.

Conservators rely on networks and relationships with collection managers, curators and donors to provide as much provenance and significance information as possible. Referring to secondary resources in order to 'fill in the gaps' where provenance information may be vague is also necessary. This may include brief or in-depth research of authoritative sources. Paul Conway suggests in his article *The relevance of preservation in a digital world*, that there are some fundamental differences between conserving analogue collections and digital preservation (Conway, 2007). Conway remarks that digital preservation is driven by the ongoing long-term access of a digitised item or born-digital collection, whereas preservation strategies for tangible collections are determined by the immediate needs of the item to ensure its stability and longevity (Conway 2007). While this is somewhat true, digital preservation practices could learn and borrow from conservation processes. Determining the initial needs of a digital item at the moment of acquisition is crucial.

Significance for digital items may also be determined by active use and demand, much in the same way that the significance of physical items can be influenced by trends, time and context (Conway 2007). The prevalence of digital culture in a global networked world also shifts our notions of 'active demand'. Whereas historically active demand may have been driven by researchers' needs to access a particular historical document, in the digital age active demand can be driven by a YouTube video going viral and being accessed by millions of users in the space of several hours. Just as physical items may have an inherent significance based on their provenance, so too may digital items. Both the tangible and intangible can become more or less significant based on access, use and demand (Howell 2001).

The immediacy of digital collections also demands that organisations respond in the same timely manner. This can be illustrated by the constant chase by web archivists to capture websites before they disappear. In current society, websites are an accessible point of reflection for culture and history. At least for a time, 'on the Internet nobody knows you're dead' (Neddam 2010); however, websites are also inherently timesensitive and ephemeral in nature. It is estimated that the average website exists for somewhere between 44 and 100 days (Guy 2009). For example, a significant online resource for Australia is the Sydney 2000 Olympics website. This site was only live for two weeks, and was archived by the NLA each day of this period. The archived version of the site is now publicly accessible via PANDORA (Preserving and Accessing Networked Documentary Resources Australia), the NLA's web archive. Another example of the power of an archive such as PANDORA is illustrated by the archiving of the Tasman Solo - The Trans-Tasman Solo Kayak Expedition website. Andrew McAuley attempted the first solo kayak expedition across the Tasman Sea, during which he blogged about his voyage. The site displays some of McAuley's final photos and blog entries before his untimely death. While the original blog is still available online at the time of this publication, it is expected that this will not always be so. The blog was archived with permission from McAuley's widow. McAuley's blog documents the first attempt to kayak solo across the Tasman Sea and his public legacy remains accessible beyond the life of his website.

Authentication

Authenticity of collection items is an influential factor in conservation work.

Collection custodians need to ensure that 'original' items are acquired to meet user

expectation. Users of physical collection items often have expectations, especially where rare items are concerned, that they will be 'experiencing' the original, whether via actual interaction or viewing it in a glass cabinet. Recently the NLA acquired an autographed 1950s silver gelatin print. For all intents and purposes, its provenance was purported to be genuine by the seller. When the newly purchased acquisition arrived, the image was found to be a duplicate of a scanned copy of an original photograph, which was obvious to the NLA's conservators without the need for deframing. Needless to say, the purchase was returned and refunded.

What if the photograph had been a digital file? Determining the authenticity of digital objects is equally as important. As we are all entrusted with protecting the integrity of our state and national collections, being able to determine authenticity is essential, irrespective of their physical or digital nature. A range of strategies is currently in use to ensure the authenticity of digital materials. One of the more common strategies is a fixity check, whereby an initial value is generated and, at a later date, compared to the original in order to ascertain whether any changes to the file have taken place.

Various technologies can be utilised such as checksums, message digests and digital

signatures. Technical metadata such as time stamps and digital signatures also supports authentication. Additional metadata also plays a role in an object's documentation, provenance and event tracking such as versioning of direct, migrated or transformed copies, adding to the object's provenance and audit trail.¹

The International Research on Permanent Authentic Records in Electronic Systems (INTERPARES) project has researched this issue in depth (Muir 2005). In most cases, librarians and archivists can control this issue by documenting access procedures and successive modifications to a given digital record (Muir 2005).

The way in which digital materials are displayed or presented can also affect the interpretation as well as the authenticity of a digital item. Limitations of, and dependencies on, capture, display and re-presentation software and hardware technologies are an uncontrollable factor. Considerable effort should be undertaken to ensure that the integrity and authenticity of an item are maintained. User interaction with a digital item may enhance, misrepresent, or obliterate the intended meaning (Webb 2005). It is also essential that the creator's intent as well as the context is respected and understood. Paul Conway recommends that a digital preservation policy should strive to present the digital information in the context most appropriate to the user's needs without compromising the aesthetics (Conway 2007). Conservators are

driven by a similar ethos, striving to maintain long-term access to an item without compromising its integrity or authenticity. While this is an ideal approach and should be followed in the majority of cases, it may not be possible for all digital materials.

Accessibility and preservation

It is understood universally that preservation strategies should enhance rather than compromise access to collections. This is also applicable in the digital realm. Before access to digital content can be provided to users, active management and ongoing preservation of digital content is necessary. Issues and risks affecting digital materials include the short lifespan of physical format carriers and the rapid obsolescence of the software and supporting hardware systems required for interpretation (Webb 2005).

Two well-known approaches discussed in digital preservation literature are emulation and migration as strategies for ensuring long-term preservation (Granger 2000). Neither approach is flawless (Granger 2000). Emulation essentially acts as a time capsule (Granger 2000). Emulation involves packaging together the digital content and the supporting software and operating system of the computer. When future users wish to access the digital content, not only do they gain access to the digital content, they will experience the digital content within the context in which it was created. This is the most common strategy used for old arcade video games. As a preservation strategy it is highly time and resource intensive.

Migration is a feasible current solution for collecting institutions (Wheatley 2001; Granger 2000). Digital content intended for specific software and hardware will at some point not be accessible using these tools. Migration is used as a strategy to continuously create new versions of the content suitable for the tools of the day. There are of course risks involved, including various processes that change files from one encoding or format to another that may result in information loss. This of course affects the integrity and authenticity of a digital item (Wheatley 2001). Migration could be likened to the of a game of Chinese whispers where the original phrase, depending on the accuracy of replication by the participants, becomes more or less distorted each time it is passed from one individual to another. Despite the risks, migration might be the only available strategy for some organisations given the limited resources and budget restraints. As active digital preservation management strategies are still in their infancy in Australia, it is imperative that organisations

retain the 'original' digital file as future digital preservation attempts and approaches may be necessary.

Responsible stewardship

While there are several approaches to actively managing, preserving and providing access to digital materials, there is no standardised approach. The Consultative Committee for Space Data Systems Open Archival Information System Reference Model (OAIS) developed by National Aeronautics and Space Administration (NASA) is aimed at addressing this problem (Muir 2005). It is essentially a conceptual road map for managing and preserving digital information using a standardised method (Lavioe, 2000). Colin Webb, former head of Web Archiving and Digital Preservation, has commented that the OAIS model has helped focus the discussions about digital archives management into a common language (Webb 2005). Fine tuning of overall concepts of digital content management are required within institutions and across the cultural heritage sector.

Responsible stewardship requires an institution's acknowledgment of the resources and means within which it must operate in order to do what is in the best interest of the collection and its users. The Koninklijke Bibliotheek (KB) – the National Library of the Netherlands – has published extensively on its method of ingesting quantifiable file formats through a risk assessment, which has been developed in-house. The KB created an e-depot, which currently archives over ten million electronic publications and journals. To ensure a systematic method of ingesting born-digital collection material into the e-depot, which initially only supported Portable Document Format (PDF) files (Rog and Van Wijk 2008); the KB implemented a quantifiable-formats approach. In accordance with institutional requirements for the e-depot to also bank increasingly on various file types, the KB developed several selection criteria to inform donors of sustainable formats that could be supported in the short and longterm for the library (Rog and Van Wijk 2008). Selection criteria such as openness and format availability, adoption, complexity, technical protection mechanisms, selfdocumentation such as metadata, robustness of format and physical carrier, and software and hardware dependencies act as a guide for the acquisition process before digital material is accepted and ingested into the collection (Rog and Van Wijk 2008). The benefits to the institution that implements this policy are obvious. They have a considerable level of control over the type of digital material collected currently and

into the future. There are, however, shortcomings with this approach. Such a regimented approach to digital preservation would be incredibly resource-intensive to apply retrospectively. Institutions such as the NLA have already amassed copious amounts of digital content in a wide range of file formats and on various physical format carriers. Currently the NLA is in the process of developing its digital collections and preservation policy. This is necessarily driven by important collection acquisition principles, regardless of whether the content is physical or digital. Current literature attempts to formulate what is needed for a successful digital preservation policy within any institution. The literature advises that any digitisation activity should be tied to the strategic directions of a collecting institution (Henry, 2000). It is also recommended that guidelines should define what aspects of a digital collection should be preserved and why (Henry, 2000). A successful digital preservation policy would also address the preservation needs of digital items created by an institution itself, such as photographs from digitisation programs. These recommendations would ideally be underpinned by a mandate, which would indicate the level of commitment to digital preservation activities (Henry, 2000; Keene, 2001; Conway, 2007). As conservators already work in accordance with statements of preservation intent, collection development policies and work plans, these concepts are already embedded in institutional collection maintenance programs. If a list of suitable file formats was formulated it would have to be revised annually in order to keep up with the constant evolution of the digital domain and the rapid emergence of new file formats, CODECs and devices. For such a policy to be implemented, curatorial decisions would need to be made on what would be collected in the first place. Factors such as significance, user demand, complexity of format and available resources (including budgetary limitations) would all impact on the acquisition process of digital collections.

At the National Library of Australia

The NLA's current strategic directions, *The National Library of Australia's Strategic Directions 2012 to 2014*, are underpinned by acknowledgement of the rapid growth of digital culture, and the agency's responsibility to collect, manage and preserve this rapidly growing field. The latest strategic directions acknowledge the challenges that digital collecting brings. Current abilities to manage and preserve digital collections must evolve in order to become the desired 'National Digital Library of Australia'.

The NLA's Preservation Services already conduct care and handling sessions on a regular basis for staff who work directly with collection materials. The Digital Preservation team has recently begun similar training sessions so that collecting staff can understand, assess, select and process digital collection materials. Both sections are committed to raising awareness of the importance of preserving physical and digital collections within the NLA.

Significant born-digital materials acquired by the NLA, either individually (such as files received as email attachments) or as a group on a physical format carrier, such as a CD-ROM, are ingested into the NLA's Prometheus workflow system. The Digital Preservation team has developed Mediapedia, an online knowledge-base resource for identifying physical and digital carriers and their associated dependencies.

Mediapedia is an extensive resource that encompasses a range of media formats, including various types of photographs, ¼" audio tape, 35mm cine film, Zip disks, and CompactFlash cards. Mediapedia is an important step towards sharing knowledge and assisting the broader community with digital preservation issues.

While the NLA is well regarded within the international community, it has by no means found the solution to digital collecting and all the associated challenges. This paper intends to progress thinking surrounding preservation as a whole, aggregating knowledge from conservation and digital preservation in order to build awareness within both communities that there are more commonalities than differences. The desire is to see the commencement of a more encompassing and holistic dialogue about this diverse area of digital culture that is part of our everyday.

Conclusion

Rather than limiting our skills and expertise as conservators, the digital question brings with it some challenges as well as many options and a great deal of opportunity. There are several major challenges with digital collecting such as potentially unstable physical carriers, evolving and increasingly complex file formats and rapid obsolescence of software and hardware devices. Digital preservation is a complex field with many more problems to be faced along the way.

We are at the early stages of discovering and understanding a new age of cultural identity created, captured, mediated and delivered by digital means. Active digital preservation is yet to become mainstream practice in many cultural organisations; it is an ideal time to impart robust traditional preservation and conservation concepts and

strategies to this burgeoning field. This opportunity presents new ways in which to share conservators' knowledge and expertise, enabling better positioning by cultural institutions to undertake new approaches for collecting, recording and protecting cultural identity and heritage in all of its forms.

Acknowledgements

Particular thanks are owed to Colin Webb for his initial encouragement and discussion surrounding the themes of this paper. Monica Omodei, David Pearson and Jonathon McCabe are also thanked by the authors for their interest, time and support throughout the research phase of the paper. The authors of this paper acknowledge and appreciate the support and assistance offered by both the Digital Preservation team and the Conservation team at the National Library of Australia.

Endnote

¹ Evolving metadata and auditing standards and methodologies such as PREMIS (http://www.loc.gov/standards/premis/), TRAC (Trustworthy Digital Repositories Audit and Certification http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying-0) and DRAMBORA (Digital Repository Audit Method Based on Risk Assessment http://www.repositoryaudit.eu/) provide strategies to consider in any digital collection stewardship planning.

References

Adam, S, 2010, *Preserving authenticity in the digital age*, School of Library and Information Science, San Jose State University, USA, in *Library Hi Tech*, vol. 28, iss. 4, pp. 595-604, viewed 10 July 2012,

http://www.emeraldinsight.com/journals.htm?issn=0737-8831&volume=28&issue=4&articleid=1896431&show=html

Conway, P, 2007, *Preservation leaflet: The relevance of preservation in a digital worl d*, Northeast Document Conservation Center, Andover, MA, viewed 23 April 2012, http://www.nedcc.org/resources/leaflets/6Reformatting/04RelevanceOfPreservation.php>

Day, Michael, 2003 'Collecting and Preserving the World Wide Web: a feasibility study undertaken for the JISC and Welcome Trust', v.1.0, 25th February 2003, http://www.jisc.ac.uk/uploaded_documents/archiving_feasibility.pdf>

Del Pozo, N, Stawowczyk, A, & Pearson, D, 2010, 'Land of the lost: A discussion of what can be preserved through digital preservation', *Library Hi Tech*, vol. 28, iss. 2, pp. 290-300, viewed 10 July 2012,

http://www.emeraldinsight.com/journals.htm?issn=0737-8831&volume=28&issue=2&articleid=1864755&show=html

Granger, S, 2000, 'Emulation as a digital preservation strategy', *D-Lib Magazine*, October, vol. 6, no. 10, viewed 5 March 2012, http://www.dlib.org/dlib/october00/granger/10granger.html>

Henry, W, 2000 'Digital Preservation: A Conservator's Perspective', Stanford Universities Libraries

Howell, A, 2001, 'Preserving information in a digital age: What's the difference?', *The Paper Conservator*, vol. 25, pp. 133-149.

Keene, S, 2001, 'Preserving digital materials: Confronting tomorrow's problems today', *The Conservator*, vol. 26, pp. 93–99, viewed 3 March 2012, http://www.suzannekeene.info/articles/digipres.htm

Lavoie, B, 2000, 'Meeting the challenges of digital preservation: The OAIS Reference Model', *OCLC Newsletter*, no. 243, pp. 26-30, viewed 10 July 2012, http://www.oclc.org/research/publications/archive/2000/lavoie/>

McAuley, A, 2007, 'Tasman Solo - The Trans-Tasman Solo Kayak Expedition', *PANDORA Australia's Web Archive*, viewed 10 July 2012, http://pandora.nla.gov.au/pan/117106/20100305-1023/www.andrewmcauley.com/index.html

Muir, A, 2005, 'Issues in the long-term management of digital material', in John Feather (ed.), *Managing preservation for libraries and archives: Current practice and future developments*, Ashgate Publishing Ltd, United Kingdom.

Neddam, M, 2010, 'Zen and the art of database maintenance', in *Sustainable archiving of born-digital cultural content*, Archive 2020, Virtueel Platform, viewed 23 April 2012, http://issuu.com/virtueelplatform/docs/archive2020_def_single_page>

Rog, J, & Van Wijk, C, 2008, *Evaluating file formats for long-term preservation*, National Library of Netherlands, Netherlands.

Rothenberg, J, 1998, Avoiding technological quicksand: Finding a viable technical foundation for digital preservation, viewed 2 February 2012, http://www.clir.org/pubs/reports/rothenberg/contents.html>

Russell, R, & Winkworth, K, 2010, Significance 2.0: A guide to assessing the significance of collections, Collections Council of Australia, Australia, viewed 10 July 2012, http://www.environment.gov.au/heritage/publications/significance2-0/

Sydney Organising Committee for the Olympic Games, 2000, 'Sydney 2000: Official site of the Sydney 2000 Olympic Games, 15 September – 1 October 2000', *PANDORA Australia's Web Archive*, viewed 10 July 2012, http://pandora.nla.gov.au/nph-wb/20001002130000/http://www.olympics.com/eng/index.html

UNESCO: Draft Charter on the Preservation of the Digital Heritage, 2003, viewed 23 April 2012, <a href="http://portal.unesco.org/en/ev.php-url_location-loca

Webb, C, 2000, 'The likely impact(s) of digital technology on conservation/preservation in Australia', *National Newsletter of the Australian Institute for Conservation of Cultural Materials*, no. 75, pp 3-5.

Webb, C, 2005, 'The malleability of fire: preserving digital information', in John Feather (ed.), *Managing preservation for libraries and archives: current practice and future developments*, Ashgate Publishing Ltd, United Kingdom.

Wheatley, P, 2001, 'Migration: A Camileon discussion paper', *Ariadne*, iss. 29, viewed 23 April 2012, http://www.ariadne.ac.uk/issue29/>

Authors' biographies

Douglas Elford is Policy and Development Officer in Digital Preservation at the National Library of Australia. Douglas has worked on the Prometheus project, including the development of the proposed sub-profile for mixed media as part of the Australian METS profile. Douglas has previously worked as the Digital Asset Services Coordinator for the National Museum of Australia and as Head of Imaging, New Media and the Web at the Western Australian Museum. Douglas has also been a commercial professional photographer in the private sector, both in Australia and overseas.

Lisa Jeong-Reuss has a Bachelor of Arts (Fine Arts) from ANU and Bachelor of Applied Science in Conservation from UC. Since graduating Lisa has worked at the National Archives of Australia (Sydney and Canberra), Australian National Maritime Museum, Australian War Memorial, and Art and Archival (private practice). Lisa is currently employed at the National Library of Australia as a paper conservator. During this time she has worked on a variety of paper-based materials and has notably worked closely with the Digitisation and Photography section in the library.

Somaya Langley has a background in the arts and culture, including sound, media arts, broadcasting, digital libraries, festivals and events. Somaya is currently a Digital Preservation Specialist at the <u>National Library of Australia</u> and has also worked for organisations including <u>ABC Classic FM</u> and the <u>Australian Music Centre</u>. Somaya has consulted to the <u>Australia Council for the Arts</u> to produce a report on <u>Archives in</u>

<u>the digital era</u>. Somaya was Production Manager of the <u>2010 International Society of Contemporary Music World New Music Days</u> festival, Co-Curator of <u>Transit Lounge</u> (a partner project of Berlin's <u>transmediale</u> festival) and a Co-Director of the 2008 and 2009 <u>Electrofringe</u> festivals.

Melanie Wilkinson has a Bachelor of Visual Communications / Graphic Design from the University of South Australia, a Post-Graduate Diploma in Art History from the University of Adelaide, and a Masters of Cultural Materials Conservation from the University of Melbourne. Melanie recently joined the Preservation Services unit as a paper conservator at the National Library of Australia after completing a 9-month graduate recruitment program and Diploma of Government. Melanie's interest in digitisation as a preservation solution stems from her thesis research during her time at the University of Melbourne.

Contact details

DOUGLAS ELFORD
Policy and Development Officer
National Library of Australia
Email: delford@nla.gov.au

LISA JEONG-REUSS Preservation Officer National Library of Australia Email: <u>lieong@nla.gov.au</u>

SOMAYA LANGLEY
Digital Preservation Specialist
National Library of Australia
Email: slangley@nla.gov.au

MELANIE WILKINSON
Paper Conservator
National Library of Australia
Email: mwilkinson@nla.gov.au

13