National Library of New Zealand, Digital Preservation and the Role of UNESCO

Steve Knight
Programme Director Preservation Research and Consultancy, National Library of New Zealand

Abstract
In 2003, UNESCO published its Guidelines for the Preservation of Digital Heritage and adopted its Charter on the Preservation of Digital Heritage. Also in 2003, The National Library of New Zealand Act was passed providing the mandate for the Library to begin collecting and preserving New Zealand’s digital heritage in ways that will ensure current and future access. This paper will show that the 2003 UNESCO documents are still valid today, look briefly at the current state of digital preservation and outline some possible approaches for UNESCO to refresh its 2003 statements on digital preservation.

Author
Steve Knight is the Programme Director, Preservation Research and Consultancy at the National Library of New Zealand (NLNZ). From a library background Steve has experience in a range of information management disciplines, including records management and document management and the design and implementation of electronic services. Among other initiatives Steve was involved in setting up the National Digital Forum in New Zealand, the award-winning Matapihi collaboration and the National Digital Heritage Archive, NLNZ’s digital preservation programme. Steve has represented the National Library of New Zealand on the National Digital Forum Board, is a member of the Advisory Committee for the School of Information Management at Victoria University in Wellington, is on the Steering Committee for the International Internet Preservation Consortium and has represented the National Library in various European Commission activities related to the Commission’s work on digital libraries and digital preservation.
1. Introduction

At an almost identical point in time UNESCO and the National Library of New Zealand (NLNZ) formally recognise the importance of digital preservation to the safety and protection of the world’s digital heritage, UNESCO through the Charter on the Preservation of Digital Heritage and NLNZ through the revision of its legislation which provided the mandate for the Library to collect in the digital realm and which catalysed the Library’s digital preservation programme.

Earlier in 2003, UNESCO had released its Guidelines for the Preservation of Digital Heritage prepared by Colin Webb, then Director of Preservation at the National Library of Australia. In 2004 the Library launches its National Digital Heritage Archive (NDHA) project, the foundation for its digital preservation programme. The project is funded by government to the tune of NZ$24 million and has a four-year timeframe.

In this paper I will undertake to do three things:

---


1. Show the current validity of the 2003 UNESCO documents with specific reference to the statements on Responsibility and Principles
2. Briefly comment on the current digital preservation environment
3. Outline some possible approaches for UNESCO to refresh its commitment to the digital preservation domain and take a leading role in the long-term preservation of the digital heritage of all nations.

The point here is not to show convergence between UNESCO conceptualising as expressed in the Charter and Guidelines and evolving practice at the NLNZ. The point is to show how robust and comprehensive this early work from UNESCO was and how beneficial that work still is to NLNZ, as a representative organisation implementing a digital preservation programme.

2. Responsibility for digital preservation

In its discussion of responsibility the UNESCO Guidelines pose four questions to help determine an organisation’s for accepting a digital preservation responsibility:

1. Does the business of the organisation imply an existing or potential preservation obligation for any kinds of digital heritage materials? (Is the organisation required to take responsibility?)
2. Does the organisation have an interest in accepting a preservation responsibility? (Does it want to have a role?)
3. Does the organisation have, or could it acquire, the capacity to take on a preservation responsibility?
4. Is this really someone else’s responsibility?

For the National Library of New Zealand the answers to these questions are unequivocally yes, yes, yes and no. There is a natural fit between a national library and the need to ensure the long-term preservation of a nation’s digital heritage. The Guidelines also propose some possible continuums for levels of responsibility an organisation might agree to undertake.
4. Using these proposed continuums of responsibility the Library’s responsibilities for the digital heritage of the nation requires a broad programme, a long-term programme, a comprehensive programme and a fully reliable programme.

The National Library of New Zealand cannot avoid digital preservation.

### 3. Principles of digital preservation

The following table reflects the NLNZ approach to the principles of digital preservation articulated in the UNESCO Guidelines.

Twenty-two of the forty-one 41 principles are reflected upon here. The reason a subset of the principles has been chosen is that one principle is often refined further by another principle, e.g. maintaining accessibility. The purpose of this exercise is to show the overall alignment of the UNESCO Principles and the NLNZ digital preservation programme as it developed over time.

<table>
<thead>
<tr>
<th>Number</th>
<th>UNESCO Guidelines principle</th>
<th>NLNZ approach</th>
</tr>
</thead>
</table>

Using these proposed continuums of responsibility the Library’s responsibilities for the digital heritage of the nation requires a broad programme, a long-term programme, a comprehensive programme and a fully reliable programme.

The National Library of New Zealand cannot avoid digital preservation.
1. Not all digital materials need to be kept, only those that are judged to have ongoing value: these form the digital heritage.

2. The goal of the NDHA programme is to enable the National Library of New Zealand … to collect, make accessible, and preserve in perpetuity, New Zealand’s digital heritage, as defined by the Library’s current collection policy, a clear recognition that similar decisions need to be made when building digital collections as with analogue collections.

3. Digital materials cannot be said to be preserved if access is lost.

4. As above the goal of the NDHA programme was to ‘to collect, make accessible, and preserve in perpetuity’.

5. Digital preservation will only happen if organisations and individuals accept responsibility for it.

6. A national library is a natural fit for digital preservation. The National Library Act 2003 gave NLNZ the mandate to collect digital and the impetus to develop its digital preservation programme. The core issue here is to be able to transfer our trustworthiness from the analogue world to the digital world.

7. It is important to do no harm.

8. The library identified nine architectural qualities as being of particular significance to the NDHA. The highest ranked of these is Data Assurance expressed as ‘zero data loss.’ The complexity between the statement and the practical import of the statement is the challenge.

9. Acceptance of responsibility should be explicitly and responsibly declared, taking account of the likely implications for other preservation programmes and for other stakeholders.

10. Acceptance of responsibility is an explicit outcome of the National Library Act 2003. We are currently working with Archives New Zealand to leverage the NDHA to support the preservation of the digital public record in line with the Public Records Act 2005. We are also working with external organisations with regard to third party hosting and the potential for a whole-of-country approach to digital preservation.

11. Working with producers to influence the standards and practices they use, and to increase their awareness of preservation needs, are important investments.

12. The Library is very aware that there is a continuum of activities that influence the nature and extent of the digital preservation programme's ability to undertake its role. We are beginning discussions with major publishers to ensure that they are integral.
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>Digital heritage materials must be moved to a safe place where they can be controlled, protected and managed for preservation.</td>
<td>The Library has its own dedicated data centre with discrete provision for the digital preservation programme. This is currently being enhanced with movement of the NDHA to a state-of-the-art data centre located outside of the Library and built to withstand one-in-a-hundred-year events. This is necessary in a geologically active region.</td>
</tr>
<tr>
<td>17</td>
<td>Digital heritage materials must be uniquely identified, and described using appropriate metadata for resource discovery, management and preservation.</td>
<td>All objects in the NDHA receive a unique and persistent identifier (using the Handle System). As well as being uniquely identified in this manner each item is discretely described in the Library resource discovery and collection management systems and data is automatically synchronised between the collection management and digital preservation systems.</td>
</tr>
<tr>
<td>19</td>
<td>Preservation programmes should use standardised metadata schemas as they become available, for interoperability between programmes.</td>
<td>The Library’s resource discovery and collection management systems are based on MARC and ISAD-G metadata. Synchronisation between collection management and digital preservation systems is via Dublin Core. In 2003 the Library presented a revised version of its earlier work on preservation metadata including an associated XML schema. With the development of the Library’s digital preservation system Rosetta, by Ex Libris Group, the underlying data model has been created in accordance with the de facto standard PREMIS.</td>
</tr>
<tr>
<td>20</td>
<td>The links between digital objects and their metadata must be securely maintained, and</td>
<td>The NDHA programme was developed within the framework provided by NASA’s</td>
</tr>
</tbody>
</table>

---

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>the metadata must be preserved.</td>
<td>Open Archival Information System (OAIS) reference model. The Library has core enterprise systems for resource discovery and collection management. These have now been joined by Rosetta, a purpose built digital preservation system. Together these provide the links between objects and their metadata required to ensure the long-term viability of the objects over time.</td>
</tr>
<tr>
<td>21</td>
<td>Authenticity is a critical issue where digital objects are used as evidence.</td>
<td>All objects ingested into the NDHA permanent repository receive a combined hash key of MD5, CRC32, and SHA1 for use in the detection of change in their files. Regular, ongoing crawling of the permanent repository is undertaken in order to detect change in the files stored there.</td>
</tr>
<tr>
<td>22</td>
<td>Data that underlies digital objects must be safely stored and managed if there is to be any chance of re-presenting authentic objects to users.</td>
<td>Issues regarding the preservation of databases and research data are still to be formally addressed by the NDHA. This is seen as a major gap in our current capability.</td>
</tr>
<tr>
<td>24</td>
<td>Authenticity is best protected by measures that ensure the integrity of the data is not compromised, and by documentation that maintains the clear identity of the material.</td>
<td>Our digital preservation system Rosetta currently contains data about 276 separate events that occur in the system. Of these fifty are considered provenance events and are retained with their relevant objects for the long-term.</td>
</tr>
<tr>
<td>25</td>
<td>Data protection is built on the principles of system security and redundancy.</td>
<td>The NDHA defines nine core architectural qualities: Data Assurance, Security, Portability, Flexibility, Manageability and Maintainability, Scalability, Performance, Availability, Disaster Recovery. We have not yet tested how well we have met our goals in terms of achieving these qualities.</td>
</tr>
<tr>
<td>28</td>
<td>Preservation action should not be delayed until a single ‘digital preservation standard’ appears.</td>
<td>Migration and emulation are not yet well proven in the digital preservation domain, in particular the quality assurance processes for determining success of any particular transformation. Understanding of formats, tools for characterising, validating and extracting</td>
</tr>
</tbody>
</table>

---

metadata from formats are all less satisfactory than we would like. However, these are not reasons for not moving towards a functioning digital preservation programme. The NDHA has material flagged as format ‘unknown’. We expect to come back to this material when resources, tools, capability etc. allow. In the meantime it is in a managed digital preservation environment.

| 30  | It is reasonable for programmes to choose multiple strategies for preserving access, especially to diverse collections. They should consider the potential benefits of maintaining the original data streams of materials as well as any modified versions, as an insurance against the failure of still uncertain strategies. | The NDHA maintains bit streams, Preservation Masters, Modified Masters (e.g. sound recordings with noise removed) and access derivatives appropriate for the individual object. The hope is that we should be able to move seamlessly between levels of objects over time as new tools and technologies arise that enable better preservation and access processes for those objects. |
| 32  | Preservation programmes are often required to judge acceptable and unacceptable levels of loss, in terms of items, elements, and user needs. | Zero data loss is the goal. However, that may not be possible for all material at all times, if at all. Pre-conditioning (making objects preservation ready), preservation planning and action may require acceptance of loss. Within the NLANZ programme this is a curatorial decision based on the best advice possible from the digital preservation programme at the time. |
| 37  | The costs of preservation programmes are hard to estimate because they encompass so much uncertainty. | Issues regarding the costs of digital preservation at NLANZ are still to be formally addressed. The move to Infrastructure as a Service (IaaS) in the next 12 months will provide more certainty about the costs of storage in the first instance. |
| 38  | Preservation programmes may start as pilot projects but they eventually need to establish sustainable business models. | Issues relating to the business model for digital preservation have not yet been addressed by NLANZ. As we move to IaaS, 3rd party hosting and the potential for a whole of country approach to digital preservation this will become increasingly important. |
| 39  | While suitable service providers may be | It is noteworthy that commercial service |
found to carry out some functions, ultimately responsibility for achieving preservation objectives rests with preservation programmes, and with those who oversee and resource them.

| 40 | Working collaboratively is often a cost effective way to build preservation programmes with wide coverage, mutual support and the required expertise. |
| 41 | Collaboration involves costs and choices as well as potential benefits. |

providers are conspicuous by their absence and where services are starting to appear they are in the ‘softer’ areas of the preservation domain, e.g. in the area of audit and certification, but not in the development of commercial emulation services. Is it that there is not a sustainable market for such services? If that is the case how does the community respond to this?

Rosetta is one of only two commercially available digital preservation systems in the world. NLNZ decided early on to work with commercial partners (Ex Libris and Sun Microsystems). We did not have the skills in-house to build a solution. We did not want to build a bespoke solution for the National Library of New Zealand. We believe our activities should be applicable as broadly as possible across the whole digital preservation community. Consequently, development of a broad-based user community, a formal programme of community agreed enhancements and a development roadmap among other things convinced us that the commercial route was more likely to provide the continuity that we needed.

As part of the NDHA project we instituted an international Peer Review Group to provide a check on the Library’s thinking about digital preservation and also to act as an objective advisor to the Library’s vendor as to whether what we were asking for did actually represent current best practice thinking about what a digital preservation system should do. We also instituted a Cross Government Working Group from 27 government departments to help catalyse awareness of digital preservation and also to tap into specialist expertise from the wider public sector.
From this brief discussion of the responsibilities and principles delineated by the UNESCO Guidelines and how they have been responded to (albeit unwittingly) by the National Library of New Zealand, we can see the quality of the Guidelines approach and its robustness. It is worth considering in particular the foresight required to articulate some very difficult practical issues, e.g. Principles 8, 22, 25 and 28.

4. Digital Preservation

In 2007, the amount of digital information created surpassed, for the first time, the amount of storage needed to deal with it. Of course, we don’t need to store all the bits created like digital TV signals or phone-call routing information. But if we wanted to, we couldn’t.9

About the only growth rate that hasn’t gone negative since the beginning of the recession in 2008 is the creation of new digital information. People are still taking pictures, making phone calls, sending emails, blogging, and putting up videos on YouTube. Enterprises are still capturing daily transaction records and adding to their data warehouses. Governments are still requiring more information be kept and protected, forcing the migration to digital TV and taking surveillance photos of their citizens.

But what is the current state of our ability to collect this digital deluge and preserve it for the long-term?

A quick perusal of the digital preservation world today shows over fifty initiatives working on digital preservation systems, repositories, projects and advisory programmes and the provision of standards, products, tools and services (and this from a decidedly narrow, monolingual/English perspective).

When we look at how we talk about digital preservation (repositories, data archiving, digital archiving, life cycle, digital curation, data curation digital preservation) and when we look at the social and cultural issues related to digital preservation (standards, audit/certification, technical, organisational, legal, economic, education) are we presented with a clear, coherent framework for the long-term preservation of nations’ digital heritage?

It could be argued that there is currently a genuine lack of a clear, shared understanding of what digital preservation is, what it should be doing, what questions it should be answering for the future and what sort of information/data we should be providing today to allow those questions to be accurately answered in the future.

This needs to be addressed, as it is digital preservation that will ensure maximum leverage of the digital long-tail. Governments everywhere are concerned about the cost to the public and the sustainability of government information. Government data, big data, linked data, open government, open access, citizens’ rights to information, data re-use all permeate our discussions about digital information. Digital preservation needs to be at the centre of this discourse.

Concurrent with this however, as noted recently in Wired Magazine, open data is not just about empowering the empowered and open data is not an end in itself. ‘Massive data dumps and even friendly online government portals are insufficient,’ … ‘ordinary people need to know what information is available and they need the training to be conversant in it’ … ‘and if people are to have more than theoretical access to the information, it needs to be easy and cheap to use.’ ‘That means investing in the

---

kinds of organizations doing outreach, advocacy, and education in the communities least familiar with the benefits of data transparency. If we want truly open government, we still have to do the hard work of addressing basic and stubborn inequalities. However freely it flows, the data alone isn’t enough.10

In her keynote at the Aligning National Approaches to Digital Preservation conference in Tallinn last year Laura Campbell broached the possibility of ‘an international preservation body with a focus on policy, perhaps assisted by an advisory expert group to identify what categories of digital objects are most at risk. The body could promote an international notion of collection, work on standards and tools, and maybe maintain a common index of preserved materials.’ The results of that conference have recently been published with six key strategies for alignment described – legal, organizational, standards, technical, economic and education.11

Maybe here lies the solution to the current fractured state of the digital preservation domain mentioned above, and maybe also a role for UNESCO.

5. Facets of the web

While the current explosion of digital materials as the result of global digitisation programmes creates the need for a supporting digital preservation infrastructure this is not where the primary risk for digital resides. The real risk is embodied in digitally born materials, those that have no analogue equivalent. And this risk is most clearly manifested by the Internet and whatever successors may emerge over time.

This is reflected in the Guidelines which note that ‘this digital heritage is likely to become more important and more widespread over time’ and ‘new forms of expression and communication have emerged that did not exist previously. The Internet is one vast example of this phenomenon.’

I want now to note a few attributes of the web that together comprise a unique challenge to the digital preservation programme and which broaden the conversation regarding the impact of digital both globally and across all of our lives.

These attributes include:

The potential to leave us with more and richer histories of moments in time - social history, minorities, ethnic communities - the nature and extent of heritage and history have the potential to be opened out.

The historian Henry Steel Commager stated about newspapers that “this is what really happened, reported by a free press to a free people. It is the raw material of history; it is the story of our own times.” 12 Increasingly this will be the role of the Internet.

The range of people engaging in digital culture far outstrips the range of people engaging in print culture. The nature of those participating in digital culture is substantially more diverse. More histories will be made available to us over time. These histories would not have left a trace in print culture.

The Guidelines note that ‘using computers and related tools, humans are creating and sharing digital resources - information, creative expression, ideas and knowledge ... that they value and want to share with others over time as well as across space.’ They also note that ‘increasingly this is a heritage that documents the actions of governments, the results of scientific research, the debate of ideas, the aspirations and imagination of communities, the histories of the current and coming world.’ Almost everyone, everywhere is online in one form or another, or will be.

Social media is about direct, instantaneous communication and can result in the creation of social movements which may never have been recorded in the past. Our digital collecting and our preservation programme need to reflect this.

Where recommendations used to be from semi-elitist sources – the food critic, the movie critic – the range of ‘experts’ that we rely on for input across the universe of our choices has increased dramatically. What are the implications for citizenship and an active democratic process? Do we face a democratisation of recommendation / information or are we heading towards a Tower of Babel? Will we become more engaged or will we become more passive?

The Guidelines note that ‘definitions of heritage need to be seen in context’, that ‘heritage value may also be based on what is important at a group or community level’, that ‘heritage materials can exist well beyond the limits suggested by national legislation or international conventions’ and, finally, that ‘anything that is considered important enough to be passed to the future can be considered to have heritage value of some kind.’

6. A Role for UNESCO?

In the above I have tried to show that the UNESCO Charter and Guidelines were exemplary documents of their time using the sections on Responsibilities and principles as a litmus. They traversed the core issues relating to the long-term safety of national digital heritage and provided a conceptual framework, which is still valid today. I have also suggested, in the section on digital preservation that the promise of these documents has not yet been met and that digital preservation practice still falls far short of the objectives that the authors of the Charter and Guidelines had envisaged.

In the following sections I want to show that UNESCO not only has a role to play in the digital preservation domain but that it has a unique role related to its global reach and objectives.

The potential to collect, preserve and make accessible a fuller expression of the cultures, heritages, histories of peoples is within our grasp. This is not to say that traditional activities related to selection and appraisal and the related assigning of value will not continue. However, it is to say that the opportunity exists in the digital sphere and within the parameters of a robust, scalable digital preservation programme to ensure that substantially more of the multiple histories of the world can be kept for the benefit of the future.

In an article for the St Louis Post-Dispatch Bill McClellan describes the agonising choice as to whether to keep a colleague and friend’s letters and papers after he died and the serendipitous path that
lead him to pass the papers on to the St. Louis Media Archives. This is not a particularly unusual occurrence in itself but how will this story be played out in the ubiquitous world of the internet? How will that ubiquity shape what history, what culture, what heritage of peoples’ is still available and accessible to the future?

Why should UNESCO be a part of the global digital preservation programme? It is not an accident that national libraries, archives, museums are called memory institutions and the history of deliberate destruction of libraries is as long as the history of print culture itself.

- Destruction of the Library of Alexandria – date uncertain
- Destruction of scientific and philosophical library in Cordoba – 10th century
- Destruction of the Corvina Library in Buda – 1526
- Destruction of the Fatimid Library in Cairo – 1806
- Destruction of the libraries and archives of the Maya – Spanish Colonialists
- Destruction of the libraries and archives of the Aztecs – Spanish colonialists
- Destruction of the National and University Library of Bosnia and Herzegovina – 1992
- Destruction of the Abkhazian Research Institute of History, Language and Literature – 1992
- Mayor of Orange, France removing material deemed to be not truly French in support of far right National Front party – 1996
- Patriot Act, United States requires libraries to hand over details of their users – 2001
- Burning Harry Potter in New Mexico – 2001
- Destruction of National Library of Iraq including books that survived the sacking by the Mongols in 1258 when the waters of the Tigris were said to have run black with ink – 2003
- Court ordered burning of books in Cuba - 2005
- Dove World Outreach Centre, Florida – 2010, in the end refrains from burning the Koran.

What is the impulse behind these acts? What is it that is feared? If it is not the memory held within these institutions? This is why digital preservation matters and this is why UNESCO needs to be at the forefront of moves towards a global approach to digital preservation.

7. A Goal for UNESCO?

Parts of a solution should include:

- We need to be more purposive about weaving digital preservation into the wider strategic approach to our digital activities;
- We need to engage more methodically with the increasing quantity and complexity of digital materials going forwards;
- We need secure, stable, agreed relationships with large institutional creators (e.g. newspaper publishers), academic and private research producers etc.;

---

• We need to move from short term project funding to ongoing sustainable funding recognising the ongoing-ness of digital preservation;
• We need to engage with the full spectrum of national and international stakeholders to make this work;
• We need to make explicit the long-tail implications of digital preservation.

From a UNESCO point of view the following quote from Marcus Garvey is characteristic of why digital preservation is important:

“A people without the knowledge of their past ... is like a tree without roots.”

From a national library point of view the following quote is characteristic of why digital preservation is important:

“A National Library is a place where a nation nourishes its memory and exerts its imagination – where it connects with its past and invents its future.”

But even national libraries are not immune to the chill effects of government priorities, hostile economic requirements, and a lack of understanding that a nation’s self-worth, identity etc. are supported by national libraries. While this is one person’s view of a very heated current debate it is a salutary reminder that offices of culture such as national libraries and national archives do not necessarily have a privileged position to protect them from the prevailing winds. The Guidelines state that ‘making sure this burgeoning digital heritage remains available is thus a global issue relevant to all countries and communities.’ In looking to refresh the excellent work of the 2003 documents, a larger goal for UNESCO could be:

_to see the vision of the Charter embedded in the national legislations of its member states._

In this way UNESCO would be fulfilling its objectives for digital preservation as stated in Article 12 of the Charter and also reinforcing Article 19 of the Universal Declaration of Human Rights which states that ‘everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.’

---
