Digital Preservation and the Role of UNESCO: 10 Years On

The Memory of the World in the Digital age: Digitization and Preservation

26-28 September 2012, Vancouver, British Columbia, Canada

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

• The UNESCO Charter and Guidelines

• The National Digital Heritage Archive at NLNZ and Rosetta

• A quick look at the current state of digital preservation

• 2000 years of ‘Why UNESCO’

• The Web and digital preservation

• A Role for UNESCO in digital preservation
In 2007, the amount of digital information created in a year surpassed, for the first time, the amount of storage to deal with it.

Of course we don't need to store all the bits created - like digital TV signals, phone-call routing information, or old email spam.

But if we wanted to, we couldn't.
The UNESCO Charter and Guidelines
<table>
<thead>
<tr>
<th><strong>UNESCO Charter on the Preservation of Digital Heritage 2003.</strong></th>
<th><strong>National Library of New Zealand (Te Puna Mātauranga o Aotearoa) Act 2003</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>The General Conference</td>
<td>The purpose of the National Library is to enrich the cultural and economic life of New Zealand and its interchanges with other nations by, as appropriate, collecting, preserving, and protecting documents, particularly those relating to New Zealand, and making them accessible for all the people of New Zealand, in a manner consistent with their status as documentary heritage and taonga;</td>
</tr>
<tr>
<td><strong>Considering</strong> that the disappearance of heritage in whatever form constitutes an impoverishment of the heritage of all nations,</td>
<td>For the purposes of carrying out his or her duties, the National Librarian and any employee, contractor, or agent of the chief executive may possess, copy, store in electronic form (whether offline or online), and use any copy of a deposited document.</td>
</tr>
<tr>
<td><strong>Recognising</strong> that such resources of information and creative expression are increasingly produced, distributed, accessed and maintained in digital form, creating a new legacy – the digital heritage,</td>
<td><strong>Understanding</strong> that this digital heritage is at risk of being lost and that its preservation for the benefit of present and future generations is an urgent issue of worldwide concern, <strong>Proclaims</strong> the following principles and <strong>adopts</strong> the present Charter.</td>
</tr>
<tr>
<td><strong>UNESCO</strong> and <strong>National Library of New Zealand</strong></td>
<td><strong>Te Puna Mātauranga o Aotearoa</strong></td>
</tr>
</tbody>
</table>
Using these proposed continuums the NLNZ’s responsibility for the digital heritage of the nation requires a broad programme, a long-term programme, a comprehensive programme and a fully reliable programme.
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.
<table>
<thead>
<tr>
<th>No.</th>
<th>UNESCO Guidelines principle</th>
<th>NLNZ approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Not all digital materials need to be kept, only those that are judged to have ongoing value: these form the digital heritage.</td>
<td>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.</td>
</tr>
<tr>
<td>8</td>
<td>It is important to do no harm.</td>
<td>The NDHA defines nine core architectural qualities: Data Assurance, Security, Portability, Flexibility, Manageability and Maintainability, Scalability, Performance, Availability, Disaster Recovery. 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.</td>
</tr>
<tr>
<td>20</td>
<td>The links between digital objects and their metadata must be securely maintained, and the metadata must be preserved.</td>
<td>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>
</tbody>
</table>
## UNESCO Principles and NLNZ 2

<table>
<thead>
<tr>
<th>No.</th>
<th>UNESCO Guidelines principle</th>
<th>NLNZ approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>Authenticity is a critical issue where digital objects are used as evidence.</td>
<td>All objects ingested into Rosetta 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. 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>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 a major gap in our current capability.</td>
</tr>
<tr>
<td>30</td>
<td>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.</td>
<td>The NDHA maintains bit streams, Preservation Masters, Modified Masters (eg sound recordings with noise removed) and access derivatives appropriate for the individual object. 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.</td>
</tr>
</tbody>
</table>
### UNESCO Principles and NLNZ 3

<table>
<thead>
<tr>
<th>No.</th>
<th>UNESCO Guidelines principle</th>
<th>NLNZ approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>39</td>
<td>While suitable service providers may be found to carry out some functions, ultimately responsibility for achieving preservation objectives rests with preservation programmes, and with those who oversee and resource them.</td>
<td>It is noteworthy that commercial service providers are conspicuous by their absence and where services are starting to appear they are in the ‘softer’ areas of the preservation domain, eg in the area of audit and certification, but not in the development of commercial emulation services. <strong>Is it that there is not a sustainable market for such services? If that is the case how does the community respond to this?</strong></td>
</tr>
<tr>
<td>40</td>
<td>Working collaboratively is often a cost effective way to build preservation programmes with wide coverage, mutual support and the required expertise.</td>
<td>Rosetta is one of two commercially available digital preservation systems. NLNZ decided early 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. <strong>We believe our activities should be applicable as broadly as possible across the whole digital preservation community.</strong> 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.</td>
</tr>
</tbody>
</table>
The National Digital Heritage Archive at NLNZ and Rosetta
The NDHA Programme is a successful partnership with our software and hardware vendors resulting in one of the most fully featured digital preservation systems available today.

Developing a Digital Preservation System - Rosetta

Design & Build

Sun Centre of Excellence

ExLibris
The bridge to knowledge

Sun Microsystems

The National Library of New Zealand

Te Puna Mātauranga o Aotearoa

Internal Affairs
Te Tari Taiwhenua

Te Puna Mātauranga o Aotearoa
National Library of New Zealand
Rosetta – End to End Preservation Management

From producer management ➔ workflow automation ➔ delivery, audit trails & reporting ➔ format registry, preservation risk management, planning and action

- User management
- Producer management
- Deposit 1
- Deposit 2
- Validation stack
- Intellectual Entity (IE) data model
- Submission Information Package (SIP) submission
- SIP processing
- Deposit registration
- Technical analyst
- Workbench
- Consolidated appraisal workbench
- Rossetta transformers
- Deposit Application Programme Interface (API)
- Audit & provenance
- Process management
- User management API
- Permanent repository
- Format Registry
- Preservation planning
- Delivery
- Meditor
- Reports
- Back office configuration
<table>
<thead>
<tr>
<th>Requirements</th>
<th>Supported</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assure data accuracy</td>
<td>Yes</td>
</tr>
<tr>
<td>Assure data completeness</td>
<td>Yes</td>
</tr>
<tr>
<td>Data must be authentic and reliable</td>
<td>Yes</td>
</tr>
<tr>
<td>Data must be secured</td>
<td>Yes</td>
</tr>
<tr>
<td>Data must be available when needed</td>
<td>Yes</td>
</tr>
<tr>
<td>Data must be identifiable – retrievable</td>
<td>Yes</td>
</tr>
<tr>
<td>Data is viewable</td>
<td>Partial</td>
</tr>
<tr>
<td>Compliant with legal and ethical requirements</td>
<td>Yes</td>
</tr>
<tr>
<td>Sharing &amp; reusing - ready to share with broader research community</td>
<td>Partial</td>
</tr>
<tr>
<td>Storage requirements (keep the relation for non-digital data or massive amount of data that requires special treatment)</td>
<td>Partial</td>
</tr>
<tr>
<td>Support different retention periods</td>
<td>Yes</td>
</tr>
<tr>
<td>Metadata requirements (support community standards)</td>
<td>Partial</td>
</tr>
<tr>
<td>Attaching documentation of the research</td>
<td>Yes</td>
</tr>
<tr>
<td>Support the institution policy (file type, storage used, retention period)</td>
<td>Yes</td>
</tr>
<tr>
<td>Preservation planning and risk analysis</td>
<td>Yes</td>
</tr>
<tr>
<td>Allow easy deposits for researchers</td>
<td>Partial</td>
</tr>
</tbody>
</table>
Committed to Rosetta as an enterprise system

- Upgrade process
- Backwards compatibility
- Significant investment in regression suite
- Test environments

- Improved usability reduces training cost

Robustness, stability, scalability
Ongoing development of Rosetta

2.x

Preservation Planning
Risk Management
Format Library Knowledge Base
Publishing Module
Plug-in Manager

Q2 2011 Q3 2011 Q4 2011 Q1 2012 Q2 2012 Q3 2012 Q4 2012 Q1 2013

3.0

Collections Support
Format Library Improvements
New Search Mechanism (SOLR)
DROID 6 Format ID Tool

3.1

Next Generation Delivery:
• Video Streaming
• Book Readers and more...
HTTP Interface for Storage
More DAM Features

Q2 2011
Enhancing the Viewing Experience

- Internet Archive Book Reader
- Open source web based document viewer
- Flexpaper
- Library of Congress Newspaper viewer
- JW Player 5
- IIP Image Viewer
- Flash and HTML5 video player – mobile support
Note that it is important to know what we are counting and for what purpose.

Numbers as 1 September 2012

1 preservation system
contains
1,041,101 intellectual entities
made up of
7.3m master files
taking up
80 Terabytes
represented in
100 different formats

this includes:
13,123 websites, which themselves are comprised of c. 80m files

Note that it is important to know what we are counting and for what purpose.
A quick look at the current state of digital preservation
What do we mean when we talk about digital preservation?

Language issues:
- Repositories
- Data archiving
- Digital archiving
- Life cycle
- Digital curation
- Data curation
- Digital preservation
- Standards
- Certification/Audit

Social and cultural issues:
- Technical
- Organisational
- Standards
- Legal
- Educational
- Economic

Is the OAIS model still relevant or do we hold to it instead of developing more granular standards for digital preservation? How do we implement PREMIS?

We need clarity and certainty about what we mean when we say digital preservation.

Is this something that UNESCO can facilitate?
<table>
<thead>
<tr>
<th>Digital Preservation Systems</th>
<th>Distributed &amp; Managed Storage</th>
<th>Standards, Products, Tools &amp; Services</th>
<th>Projects, Advisory, Community</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digital Forever (ExLibris Rosetta)</td>
<td>iRODS</td>
<td>Atos Origin</td>
<td>iPres 02010</td>
</tr>
<tr>
<td>Oracle/MySQL</td>
<td>LOCKSS</td>
<td>CDL California Digital Library</td>
<td>Micro services</td>
</tr>
<tr>
<td>JHOVE2</td>
<td>Archived In</td>
<td>The technical registry PRONOM</td>
<td>DPC Digital Preservation Coalition</td>
</tr>
<tr>
<td>DRAMBORA</td>
<td>PORTICO</td>
<td>Open Planets Foundation</td>
<td>InterPARES Project</td>
</tr>
<tr>
<td>Digital Preservation</td>
<td>HATHI Trust</td>
<td>Oracle/MySQL</td>
<td>SPAR_BNF</td>
</tr>
<tr>
<td>DataONE</td>
<td>Fedora Commons</td>
<td>METS-Alto</td>
<td>BL DOM</td>
</tr>
<tr>
<td>Data ONE</td>
<td>Data Intensive Cyber Environments (DICE)</td>
<td>NLNZ Metadata Extract Tool</td>
<td>DAITSS</td>
</tr>
<tr>
<td>Fedora Commons</td>
<td>DURASPACE</td>
<td>Trustworthy Repositories Audit &amp; Certification (TRAC)</td>
<td>JISC</td>
</tr>
<tr>
<td>Data ONE</td>
<td>Tessella</td>
<td>METS</td>
<td>ORACLE</td>
</tr>
<tr>
<td>Internal Affairs (Te Tari Taiwhenua)</td>
<td>kopal</td>
<td>NLNZ Metadata Extract Tool</td>
<td>Te Puna Mātauranga a Aotearoa</td>
</tr>
</tbody>
</table>
2000 Years of ‘Why UNESCO’
Some Reasons Why UNESCO and Digital Preservation

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 – but the fallout continues.
UNESCO and Memory Institutions

It is not an accident that national libraries, archives, museums are called memory institutions.

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 we need a shared understanding of what digital preservation comprises.
This is why interoperability matters.
This is why we need more collaboration.

And this is why we need UNESCO
The Web and Digital Preservation
Attributes of the Web 1

The potential to leave us with more and richer histories of moments in time - social history, minorities, ethnic communities.

“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.”
Henry Steel Commager talking about newspapers in 1951 – increasingly this is the role of the Internet.

The range of people engaging 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.
Attributes of the Web 2

‘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.’
UNESCO Guidelines, 2003

‘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.’
UNESCO Guidelines, 2003

Almost everyone, everywhere is online in one form or another, or will be.
Attributes of the Web 3

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?
A Role for UNESCO in Digital Preservation – the Time is Right!
Research Data

Increasing demand for research data preservation
2012 Top Ten Trends in Academic Libraries

- Communicating value
- **Data curation**
- **Digital preservation**
- Higher education
- Information technology
- Mobile environments
- Patron driven e-book acquisition
- Scholarly communication
- Staffing
- User behaviors and expectations

... general lack of long-term planning for preservation. No strategic leadership for establishing architecture, policy, or standards for creating, accessing, and preserving digital content is likely to emerge in the near term.
Funder’s View – a UK Example

<table>
<thead>
<tr>
<th>Research Funders</th>
<th>Policy Coverage</th>
<th>Policy Specifications</th>
<th>Support Provided</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full Coverage</td>
<td>Partial Coverage</td>
<td>No Coverage</td>
</tr>
<tr>
<td>AHRC</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>BBSRC</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>CRUK</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>EPSRC</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>ESRC</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>MRC</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>NERC</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>STFC</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Wellcome Trust</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

Long-term curation

Monitoring

Guidance

Repository

Data centre

Costs

http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies
A Role for UNESCO

The UNESCO Charter and Guidelines were exemplary documents of their time.

However, the promise of these documents has not yet been met and digital preservation practice still falls far short of the objectives that the authors of the Charter and Guidelines envisaged.

UNESCO not only has a role to play in the digital preservation domain but 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.

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 alive for the benefit of the future.
“A people without the knowledge of their past … is like a tree without roots.”
Marcus Garvey, ?

“A people are as healthy and confident as the stories they tell themselves …
without stories we would go mad. Life would lose it’s moorings or orientation…”
Ben Okri

‘making sure this burgeoning digital heritage remains available is thus a global issue relevant to all countries and communities.’
UNESCO Guidelines, 2003

In looking to refresh its 2003 documents, the goal for UNESCO could be:

**to see the vision of the UNESCO Charter on the Preservation of Digital Heritage embedded in the national legislations of its member states.**