



Date : 01/07/2008

FROM THEORY TO PRACTICE: DIGITAL PRESERVATION AT THE NATIONAL LIBRARY OF NEW ZEALAND TE PUNA MĀTAURANGA O AOTEAROA¹

Steve Knight

Associate Director National Digital Library, National Library of New Zealand



Meeting:

84. Preservation and Conservation, (PAC), Information Technology, IFLA-CDNL Alliance for Bibliographic Standards (ICABS) and Law Libraries

Simultaneous

Not available

Interpretation:

WORLD LIBRARY AND INFORMATION CONGRESS: 74TH IFLA GENERAL CONFERENCE AND COUNCIL
10-14 August 2008, Québec, Canada
<http://www.ifla.org/IV/ifla74/index.htm>

The holdings of an institution such as a national library are core to our collective sense of identity. The continuing development and exploration of that identity through our present dialogue with the past, and the decisions we make now regarding our behaviours in a digital environment, will determine the resources available for that discourse in the future.

“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.”²

In New Zealand, the development of digital preservation services is being accomplished in the wider context of the National Library of New Zealand's (the National Library) legislative mandate and strategic direction. In this manner, ensuring New Zealand's digital memory is protected and safe, yet available and accessible to New Zealanders and anyone who may be interested in that material ensures that digital preservation is contributing to the wider objectives of the organization and the wider goals of government.

¹ The Māori name of the National Library means “well-spring of knowledge”.

² Ryckmans, Pierre. 1996. Perplexities of an electronically illiterate old man. *Quad-rant*, September 1996, No 329.

The Digital World

What we have seen in the last 15 years of the new digital era is the democratization of information production and access. The Library is contributing to this democratization through its vision of 'New Zealanders connected to information important to all aspects of their lives'.³

The 'tyranny of distance' no longer dictates how New Zealanders live, act, or are perceived in the world. Geography is no longer the key determinant. Increasingly, our position in the world is determined by how we express ourselves on the web, by how our businesses, our scientists, innovators, entrepreneurs, our culture, and heritage are visible and accessible on the Internet.

For researchers, students, educators, family historians, academics and all our citizens interested in the pursuit of knowledge and information who use our services, there has been a paradigm shift in their expectations of how that knowledge and information should be made available to them, and those expectations are centred on the web. This shift is of as much moment to the scholarly community and the owners of the ubiquitous digital camera as it is to libraries.⁴

This change has had no less an impact on the business of national libraries. In fact, it can be easily argued that there is nothing more important to national libraries today than how they go about responding to the changes in the wider society wrought by new technologies.

The relevance and viability of national libraries may be determined by their ability to respond to these changing expectations of our customers. It is imperative that going forward we are clear about what services we deliver, how we deliver them, and that we resource them appropriately. Otherwise, there is a very real risk that national libraries will cease to be relevant now and into the future and that one of the key pieces of a nation's information infrastructure will not have a part to play in an increasingly globalised information market.

National libraries are the trusted repositories for a nation's memory and documentary heritage and it may be that the key challenge of the digital era for national libraries will lie in ensuring the ongoing sustainability of our digital resources.

This paper canvasses some of the organisational, primarily non-technological, issues that we will need to address in order to successfully embed digital preservation as an ongoing component of our core business.

This is not to underestimate the technological issues at the heart of digital preservation, but is an attempt to ensure that social and organisational issues are not swamped by the enormity of that technological challenge both now and in the future.

In particular the paper will look at:

³ New Generation National Library Strategic Direction to 2017. <http://www.natlib.govt.nz/catalogues/library-documents/strategic-directions-to-2017>. Accessed 3 May 2008.

⁴ Lyman, P. & Varian, H. 2000. *How much information?* University of California, Berkeley. <http://www.sims.berkeley.edu/research/projects/how-much-info/>. Accessed 29 April 2008.

- digital preservation and the legislative and strategic context
- business change or organisational readiness for digital preservation management
- migration of current digital content to a preservation environment
- integration of digital preservation systems into an organisation's infrastructure
- performance measures for a digital preservation system.

Digital Preservation and the Legislative Context

The National Library of New Zealand Te Puna Mātauranga o Aotearoa Act 2003 provided the legislative mandate (including electronic legal deposit) for the National Library to incorporate digital preservation as a core component of its business activities and requires the National Library to collect and preserve digital content in ways that ensure current and future access to New Zealand's documentary heritage.

"The purpose of the National Library is to enrich the cultural and economic life of New Zealand ... 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;⁵

The National Library's digital preservation activities underpin the four strategic priorities identified in the Library's *New Generation National Library – Strategic Directions to 2017*⁶:

- Accessing New Zealand's digital memory
- Sharing our nation's stories
- Inspiring knowledge creation and economic transformation
- Enriching the users' experience.

Digital Preservation and the Strategic Context

The National Library's digital preservation activities also support the wider political context including the New Zealand government's three key priority areas:

National identity

The National Library promotes the creation, sharing and preservation of information that reflects our histories, identities, cultures, stories, language, values and beliefs.

Economic transformation

The National Library contributes to the economy through the creation and sharing of knowledge which in turn creates opportunities for new wealth building and also through its contribution to preventing costs related to crime, social inequity, and welfare dependency created through illiteracy and the lack of connectedness that comes from a lack of cultural identity.

⁵ Section 7 of the National Library of New Zealand Act 2003. <http://www.natlib.govt.nz/catalogues/library-documents/nlnz-act-03>. Accessed April 29 2008.

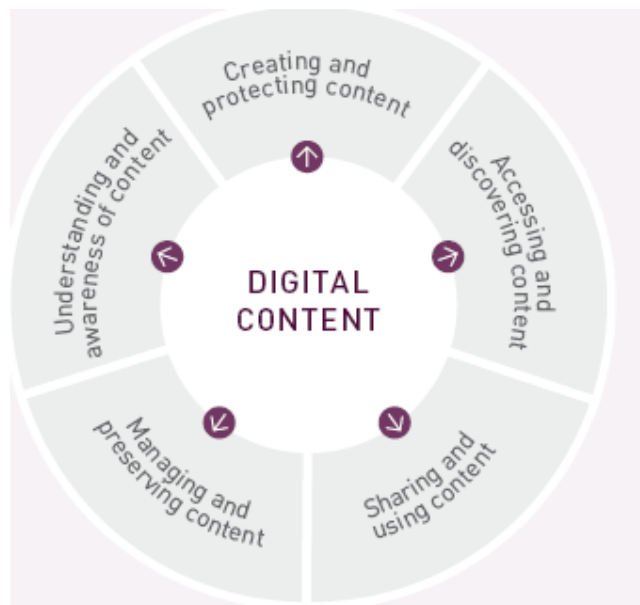
⁶ New Generation National Library Strategic Direction to 2017. 2007. <http://www.natlib.govt.nz/catalogues/library-documents/strategic-directions-to-2017>. Accessed 7 May 2008.

Families Young and Old

The National Library enables families, communities, businesses etc to connect to each other through the provision of information online and provides support for reading, learning, and literacy to help enrich people's lives.

In addition, issues of sustainability and digital preservation are central to the New Zealand government's initiative 'Creating a Digital New Zealand: New Zealand's Digital Content Strategy'. The Digital Content Strategy, development of which was led by the National Library, stresses the need to 'unlock our store of valuable content by putting it in digital form so its value can be rediscovered and renewed. As New Zealanders and end-users, we need to see ourselves on air and online, because this is the opportunity to truly promote our unique heritage, cultures and achievements, and find our place in the digital world'.⁷

Diagram 1. New Zealand's Digital Content Strategy: Five-Element Framework 2007



The above framework⁸ provides a mechanism for understanding the different dimensions of digital content and its use and applicability in the digital age. The five elements are briefly described below.

Creating and protecting content

Born-digital content is information in a new form. It needs new skills for its creation and use, provides unique opportunities for innovation and creativity and requires the means to protect it from misuse.

⁷ Creating a Digital New Zealand: New Zealand's Digital Content Strategy. 2007, p 3.
<http://www.digitalstrategy.govt.nz/Parts-of-the-Digital-Strategy/Content/New-Zealand-Digital-Content-Strategy/>.
Accessed 29 April 2008.

⁸ Creating a Digital New Zealand: New Zealand's Digital Content Strategy. 2007, p 27.
<http://www.digitalstrategy.govt.nz/Parts-of-the-Digital-Strategy/Content/New-Zealand-Digital-Content-Strategy/>.
Accessed 29 April 2008.

Accessing and discovering content

Content in digital form, whether born-digital, digitised or simply indexed digitally, competes with billions of other items of content for users' attention. It is vital to have the mechanisms for content access and discovery, such as design standards, metadata and search engine optimisation.

Sharing and using content

A key feature of the digital age is users' ability to find relevant content that they can readily use, reuse, share and repurpose and to which they can add their own dimensions.

Managing and preserving content

As different formats and devices become obsolete, digital content risks being lost much more easily than its physical equivalent. Managing and preserving content for continued use is essential if it is to survive.

Understanding and awareness of content

Digital content is altering our commonly held notions of information, knowledge and material value. We need to understand the digital content environment and its opportunities and challenges if we are to make more informed decisions, choices and investments.

Business Change; Organisational Readiness for Digital Preservation Management

Legislation and strategies on their own, however, do not provide the National Library with the resources, services and infrastructure required to support digital preservation.

At implementation level and in line with its own New Generation National Library strategy, by late 2009, the National Library's National Digital Heritage Archive (NDHA) Programme⁹ will have ensured it has the technological and organisational infrastructure to preserve and provide access to the digital heritage collections it is responsible for.

Concurrent to software development and hardware implementation, organisational readiness or 'business change' to integrate the NDHA system and digital preservation management processes is also being undertaken.

The objectives of business change are to ensure the National Library has:

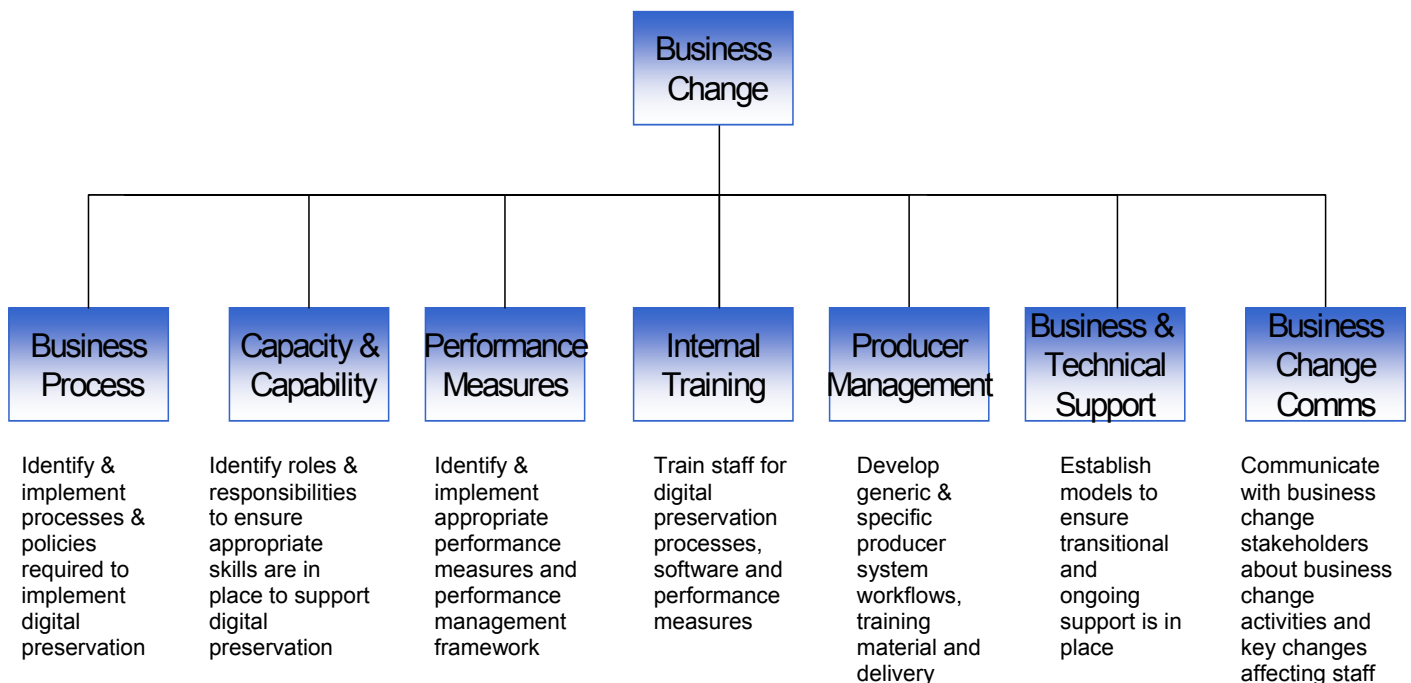
- Business processes, policies and procedures in place to support digital preservation.
- Capacity and capability to support digital preservation (including staff, environment, hardware and software).
- Performance measures and a performance management framework to support digital preservation decision-making.
- Appropriate training procedures in place for both external producers of digital content and internal staff.

⁹ <http://www.natlib.govt.nz/about-us/current-initiatives/ndha>. Accessed 7 May 2008.

- Robust workflows in place including specific workflows for producers of digital content.
- Support models in place to ensure as much knowledge transfer from the project into the business as possible and that managers, curators and team leaders are ready to assume responsibility for their components of the various workflows.

Seven discrete work streams are developing and implementing the elements required to ready the organisation for end to end digital preservation management, including the introduction of the NDHA system in late 2008.

Diagram 2. Business Change Work Streams



Only with comprehensive prior development and implementation of organisational change on this scale, can the National Library successfully integrate all aspects of digital preservation management into the business. While the software and hardware environments will make up the NDHA per se, they require interaction with the organisation's human resource as well as integration with the organization's other software environments, in particular access mechanisms, and interim Digital Object Management systems.

Integration of Digital Preservation Systems into an Organisation's Infrastructure

If the National Library's only goal were to ingest and preserve digital content in complete isolation from the other systems and processes within the National Library, then digital preservation would be a much simpler task.

However, the National Library has already made significant investments in a range of resource discovery and content management systems along with other ancillary

systems to manage and disseminate the National Library's collections. And in considering digital preservation it is clear that in many cases the management of digital content will be subject to the same systems and processes already in place, for example acquisition, description, end-user delivery and reporting.

Therefore, new digital preservation systems must integrate with other software applications the National Library uses to deliver digital library services to users. Consequently, the implementation of a digital preservation system such as the NDHA must be done within the context of the National Library's extant information and enterprise software/hardware architecture and must include, as part of the programme, integration with that infrastructure.

In other words, good architectural practice involves the consolidation and re-use of 'like typed' services wherever appropriate, and not the continual creation of independent stand alone solutions. In some cases, we have created new services specifically to support digital content, but this is only the case where no 'like typed' digital service existed.

Tools the NDHA Programme is developing to integrate the NDHA system with the National Library's collection management systems and access products have included INDIGO, an internal submission application for 'ingesting digital objects'. Other applications and integration components, for example a Web Curator Tool¹⁰, have also been developed with the understanding that those applications will be provided as open source objects for the wider digital preservation community.

Examples of some of the integration points for implementation of digital preservation within the National Library include:

Table 1: Integration points for implementation of digital preservation

Function	Activity	New or Existing
Ingest	Manual Staff Deposit	New
Ingest	External deposit	New
Ingest	Web Curator Tool	Existing
Ingest	Email deposit	New
Ingest	Digitisation	Existing
Validation	Virus Checking	Existing
Metadata Extraction	NLNZ Metadata Extract Tool	Existing
Format characterisation and validation	JHOVE, Pronom	New within National Library
Collection Management	ILS & Tapuhi	Existing
Identifiers	Handle Server	Existing
Reporting	Business Reports	Existing
Delivery	Web Delivery Modules	New
Resource Discovery	ILS, Tapuhi and other existing web based resource discovery systems	Existing
Monitoring	Systems monitoring	Existing
Enterprise services	Email, FTP, LDAP etc	Existing

¹⁰ <http://www.natlib.govt.nz/about-us/current-initiatives/past-initiatives/web-curator-tool/?searchterm=web%20curator%20tool> Accessed 8 May 2008

Migration of Current Digital Content to a Preservation Environment

The NDHA Programme developed an Object Management System (OMS) in September 2005 to ensure the National Library met its legal deposit responsibilities under the revised National Library Act prior to the establishment of the NDHA. Material currently stored in the Object Management System (OMS) will be migrated into the NDHA at the end of 2008.

The development of the OMS also gave the NDHA Programme the opportunity to test some elements of the digital preservation process in a real life scenario, for example fixity and virus checking. The OMS has developed over time as a digital repository for staff to upload a range of digital objects including:

- Published material deposited under Legal deposit.
- Digitised material from the Library's digitisation programme.
- Websites harvested as part of the Library's web archiving programme.

At present approximately 80,000 intellectual entities made up of around 280,000 files are stored in the OMS. Since September 2005, business processes and services have been created that depend on the OMS and these processes and services must be migrated to the new NDHA system along with the digital content.

While the initial migration of six terabytes may not seem like an immense amount of data to migrate, the NDHA Programme has determined that the most appropriate method of migrating the data into the new NDHA system is to deposit it in the same manner that new content would be deposited. This will have the added benefit of testing initial workflow and process configurations. This also imposes the same metadata constraints (referential integrity, data validation), validation checks (fixity verification, virus check, format identification and metadata extraction) and enrichment tasks (CMS identifier association, access derivative generation) that will be applied in a live operational setting.

While this will significantly increase the amount of processing required to ingest the data, as opposed to inserting the content and metadata records directly into the permanent repository, data integrity and security have been primary objectives since the beginning of the NDHA Programme and will be validated using this approach against the OMS data.

A significant amount of work is being performed on the data in preparation for the migration, partially due to the more stringent constraints being imposed on the data than were imposed in the ingest of the material into the OMS itself. Where necessary, data is being rectified automatically but in a lot of cases the only way to upgrade the data is to develop reports to identify affected records and update them manually.

One of the other major tasks being performed is the regeneration of a significant amount of existing access derivatives, moving from a 'just in case' JPEG generation from high resolution TIFF files, to a 'just in time' on the fly JPEG creation from an intermediate JPEG2000 access derivative. The purpose of this is to improve end-user experience when viewing the content, and to future proof delivery systems so that the National Library can provide a single point of content delivery as a service to other applications within the Library. Other access derivatives, such as archived websites,

video, audio and documents will be accommodated for within the ingest workflow of the NDHA system itself.

All of this work will give an indication of the amount of effort required to migrate the rest of the National Library's digitized content into the NDHA system. A significant amount of digitization work has been done outside of the OMS, and 12 months after the initial migration and launch of the NDHA should see those systems migrated as well. In this case an order of magnitude larger than the OMS migration. However, because this content is digitized using specific specifications and structures, and not born digital like some of the content in the OMS, the tools used to perform the migration should be somewhat simpler.

Performance Measures for a Digital Preservation System

Approximately 60 key performance measures covering key performance indicators, reporting, audit and internal ingest have been developed for the NDHA Programme to date.

Historically, the National Library has measured 'widget' style outputs for its reporting. In line with a broader trend towards measuring against outcomes, the NDHA performance measures are more targeted towards management information to support decision making related to the ongoing digital preservation programme. As well as providing the elements to be measured, this work also provides a clear statement of what type of response would be required in the case of either 'over delivery' or 'under delivery' against the measures.

A trial will be undertaken prior to the system going live with a view to introducing a culture of using performance measures, testing the current set of measures for validity and refining the measures where necessary

Table 2: Examples of digital preservation management performance measures.

Process	Success Criteria	Performance Measure	What does the measure tell us?	What actions are available in the case of over delivery?	What actions are available in the case of under delivery?	Responsible	Accountable	Consulted	Informed
Overall	The NDHA is a trusted digital repository	The NDHA has certification from the relevant internationally recognised accreditation body	What other people think of us	Review compliance against criteria and consider reducing standards.	Review compliance against criteria and consider increasing standards Consider alternative accreditation agency	NDHA Manager	Director, National Digital Library	NDHA Team	All Staff
	The integrity of objects in the digital repository is maintained	100% of objects in the permanent repository pass virus and fixity checks	NLNZ is preserving the integrity of objects in its repository	Not applicable	Review reasons for failure Identify preventative measures	NDHA Digital Preservation Analyst	NDHA Manager	DPS Administrator NDHA Preservation Policy Analyst	
Deposit material for external depositors	Digital material is deposited successfully	Number and % of successful submissions, against total attempted submissions	Whether there are problems with the deposit process	Not applicable	Review deposit process and identify problem points Review reasons for failure and take actions as appropriate	Technology Services Staff DPS Administrator	CIO Technology NDHA Manager		Curators Acquisitions & Collection Development Team Leaders Field Librarian
Deposit material for external depositors (cont)	Internal and external producer agents are satisfied with the deposit process	Producer agent satisfaction is X% in surveys	Producer satisfaction levels	Not applicable	Review areas of dissatisfaction and identify potential actions	Curators Acquisitions & Collection Development Team Leaders Field Librarian	Manager Collection Management Curator Published Collections Manager Archival Collections	Internal and external Producers	Acquisitions & Collection Development staff ATL staff who report to curators
Access material	The NDHA delivers objects as required by the resource discovery systems	Objects are delivered to resource discovery systems upon valid request	NDHA provides appropriate access to objects through resource discovery systems	Not applicable	Review availability and continuity of (and between) DPS and resource discovery systems	Technical services staff (helpdesk and availability of resource discovery systems and DPS)	CIO Technology (availability of resource discovery systems and DPS)		NDHA Manager Director National Digital Library

Process	Success Criteria	Performance Measure	What does the measure tell us?	What actions are available in the case of over delivery?	What actions are available in the case of under delivery?	Responsible	Accountable	Consulted	Informed
	Material in the NDHA is accessed by authorised users in accordance with applicable access restrictions	NDHA enforces 100% of access restrictions correctly	<p>Clients are able to access material they are permitted to access</p> <p>Clients are not able to access material they are not permitted to access</p>	Not applicable	Review reasons for failure and identify remedial actions	DPS Administrator	NDHA Manager		<p>Chief Librarian ATL</p> <p>Manager Content Services</p>
Acquire material	Material has been assessed and acquired where appropriate	Number and % of items that go through the appraisal (approval/assessment) process that are rejected, declined, and accepted, at the file, representation, and IE level, and the reasons, by appraisal type	<p>NLNZ internal and external communication of collection policies</p> <p>Technical capability of the system</p>	<p>(Means very high proportion of accepted objects)</p> <p>Review coverage of producers and content against collection policies</p> <p>Review marketing strategies and communication of collection policies (could mean we are not getting a broad enough coverage)</p>	<p>(Means high level of declined and/or rejected objects)</p> <p>High level of declined: Publicise collection policies and review marketing strategies</p> <p>High level of rejected: Review deposit guidelines available to external depositors, including screens</p> <p>Review technical capability of the system (e.g. whether format library is keeping up to step with environment)</p>	<p>Acquisitions & Collection Development Team Leaders</p> <p>Curators</p> <p>DPS Administrator</p>	<p>Manager Collection Management</p> <p>Curator Published Collections</p> <p>Manager Archival Services</p> <p>NDHA Manager</p>		<p>Acquisitions & Collection Development Staff</p> <p>ATL staff who report to curators</p> <p>NDHA staff</p>
Maintain individual objects	Objects are maintained within agreed timeframes	No. and % of maintenance activities completed within target timeframe, by maintenance type	No. and % of maintenance activities completed within target timeframe, by maintenance type	<p>Review targets</p> <p>Review resource levels</p>	<p>Identify reasons for non-performance</p> <p>Review targets</p> <p>Review resource levels</p> <p>Review staff training</p>	DPS Administrator	NDHA Manager		<p>Manager Collection Description</p> <p>Curator Published Collections</p> <p>Manager Archival Services</p> <p>Manager Collection Management</p> <p>Manager Preservation</p>

Process	Success Criteria	Performance Measure	What does the measure tell us?	What actions are available in the case of over delivery?	What actions are available in the case of under delivery?	Responsible	Accountable	Consulted	Informed
Manage exceptions	Exceptions are managed in a timely manner	No. and % of exceptions, by stage, and failure type	Identifies where there are technical problems in the system Nature of problems that occur within each stage	Not applicable	Identify reasons for failures Review deposit guidelines available to depositors Identify if any automated processes can be amended or built into the system to resolve errors	DPS Administrator	NDHA Manager		Manager Collection Description Curator Published Collections Manager Archival Services Manager Collection Management Manager Preservation

Conclusion

In New Zealand, the issue of preserving the nation's digital cultural heritage, past, present and future is addressed by legislation and central government policy. The National Library of New Zealand (Te Puna Mātauranga o Aotearoa) Act 2003 requires the National Library to collect, preserve, protect and make accessible digital collections, along with traditional paper collections, in ways that ensure current and future access to New Zealand's documentary heritage.

At implementation level and in line with its own New Generation National Library strategy, the National Library of New Zealand is ensuring it has the infrastructure, technology and organisational structure and work processes to preserve and provide access to the digital heritage collections it is responsible for.

Concurrent to software and hardware development, organisational readiness to integrate digital preservation is being strengthened. This includes business change, integration, migration, and performance measurement work streams.

There is also a need to ensure that the sustainability of digital assets over time becomes a national issue. The democratisation of production and access means that there will be a vast increase in the quantity of citizen's created content that will impact on our collecting and preservation processes. There is a need not only to provide the facilities for this content to be created but also the facilities for this content to be preserved and incorporated into a common, linked infrastructure for search, retrieval and preservation both nationally and globally.

Digital preservation will become an increasingly important component of New Zealand's knowledge infrastructure. It is the mechanism for the preservation of New Zealand's digital memory and, in the context of national libraries, must be done with a view to it becoming as unconscious in our day-to-day life as our other collection management and delivery activities are today.

Preservation without content, content without description and search and retrieval and connection mechanisms, and search mechanisms without user friendly front end access are not sufficient in themselves. One missing element degrades the value of all the others. It is important to look at digital preservation in the context of an overall strategic and holistic approach to the surfacing of our collections.

As Garrett and Waters wrote:

*' 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 maneuver effectively in a digital landscape. It is a problem of building ... the various systematic supports ... that will enable us to tame the anxieties and move our cultural records naturally and confidently into the future.'*¹¹

¹¹ Garrett, J. & Waters, D. (eds). (1996). *Preserving digital information: Report of the Task Force on Archiving of Digital Information*. Washington, DC: Commission on Preservation and Access and the Research Libraries Group. <http://www.oclc.org/programs/ourwork/past/digpresstudy/final-report.pdf>. Accessed 29 April 2008.