Securing the Future: Digital Preservation at the National Library of New Zealand

International Group of Ex Libris Users (IGeLU)

Madrid 2008

Steve Knight,
Associate Director National Digital Library
National Library of New Zealand
- Digital preservation system development with Ex Libris
- Relationship with Digitool & development of the new system
- Usability issues & resource discovery issues (for users)
- Ingest and management issues for librarians
- Deposit and ingest issues for users
- Cataloguing and metadata issues for librarians
- Relationships between items in the NDHA and other repositories/systems at the National Library
- Workflow issues for staff involved in the NDHA and other library systems
What I’m going to talk about

- The digital environment
- NLNZ drivers
  - legislation, Legal deposit
- NLNZ, Ex Libris and DPS
- Organisational issues
  - Business change
  - Migration
  - Integration
  - Capability and capacity
  - Performance measurement
- A national strategy
- Digital preservation as the sine qua non

How I hope to touch on these topics this morning
The ‘digital environment’

The digital native

Isn’t it amazing what kids know?

- Democratization of information production and access
- Citizen’s created content and will impact on our collecting and preservation processes.
- ‘New Zealanders connected to information important to all aspects of their lives’.
- Paradigm shift in client expectations of how knowledge and information should be made available to them
- Relevance and viability of national libraries determined by their ability to respond to these changing expectations
- Digital preservation will become an increasingly important component of New Zealand’s knowledge infrastructure.
- Digital preservation as part of an overall strategic and holistic approach to the surfacing of our collections.
The Problem Statement


Preserving digital information

‘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.’
Part 4, 29 (1)
In this part, unless the context otherwise requires, -

**electronic Document**, means a public document in which information is stored or displayed by means of an electronic recording device, computer, or other electronic medium, and includes an internet document

**internet Document**, means a public document that is published on the Internet, whether or not there is any restriction on access to the document; and includes the whole or part of a website

**make a copy**, in relation to an internet document, means to make a copy of the document for the purpose of storing and using it in accordance with this Part; and includes circumventing any technological protection measures that would otherwise prevent or hinder the copying, storage or use of the document
Electronic legal deposit

Why the government wants two copies of your CD

The government’s new policy – ‘steal this record’

The government declares itself in favour of the knowledge economy and is a serious promoter of New Zealand art and culture.

Yet, curiously the National Library seems determined to undermine these policies.

The National Library Act of 2003 requires all publishers in New Zealand to lodge three copies of their works with the library to build a collection of all local works. These are available to the public. So far so good. This is common practice overseas.

However, the library now proposes that the “documents” which must be lodged in their collection include electronic media such as audiotapes, videotapes, CDs, CD-ROMs and DVDs.

So if you are planning to make a CD of one of New Zealand’s choirs singing French medieval music, or a DVD of the history of Chinese families who settled in New Zealand you must lodge three copies with the library.

Even this may not seem unreasonable.

The problem is that the library will digitise this on the worldwide web for all and sundry to listen to, downloaded and copy.

The National Library seems unconcerned by the risk of copying. It ingeniously points out that normal copyright protection is available to an author, composer or publisher whose material is unlawfully copied.

Please tell me, how can a specialist music club or historical society afford to take legal action against some unknown person in France or Hong Kong who decides to download and copy a hundred CDs for sale to French or Chinese enthusiasts?

Why would anyone invest in a film in New Zealand if the film is to be digitised and posted on the web for all and sundry to download and copy? What will Creative New Zealand say when its successful productions are posted on the web for all to steal?

The Library is suggesting a time limit of three years or more before posting such material on the web. This may be OK for the latest song by some New Zealand Idol, but what about work which has a longer life?

New Zealand has produced, and continues to produce, its fair share of world-class opera singers and pianists, who will simply make sure their records are produced in jurisdictions which don’t aid and abet the theft of their intellectual property.

The assaults on private property on the foreshore and farms have been bad enough. But at least the government was stealing land from a few New Zealanders to hand over to other New Zealanders.

Why would any government want to steal the artistic works of New Zealanders to hand them over to foreigners?

I just hope I’m missing something here.

Our cyberspace publications will be available to future historians
Environment for Change

Legislation

Strategic Vision

International Community
The Other Problem Statement

Technology as the enemy?

The evolution of technology environments

Windows Vista
Word 2007

Danger Point

Windows xxx?
Word xxxx?

2004 2030
NDHA Programme

Goal
Collection, preservation and access in perpetuity

To establish the National Digital Heritage Archive to enable the National Library of New Zealand to meet its mandate to collect, make accessible, and preserve in perpetuity, New Zealand’s digital heritage, as defined by the Library’s current collection policy.
Partnership 1

We cannot do this alone

Design & Build

Sun Centre of Excellence

Ex Libris
FROM LIBRARY SYSTEMS TO INFORMATION SERVICES

Sun Microsystems
Partnership 2

Some philosophical questions

Buy or build?
Proprietary or open source?
Or are they religious questions?

- Commercial solution vs. building it yourself vs. project based company
  - User community
  - Enhancements
  - Continuity
  - Open source 80% (Jhove, Droid)
- Important to look at the required institutional outcome
- Repository solutions, digital archiving solutions and digital preservation systems are unlikely to be the same thing
Partnership 3

• Digital Preservation System (DPS)
  – generic software solution for the wider market
  – broad ranging digital preservation solution for a range of community interests

• NDHA - the NLNZ implementation of DPS
  – wider functionality and business change are required for practical digital preservation within any given institution

Not a National Library of New Zealand cul-de-sac

It is important from NLNZ perspective that the solution is not NLNZ specific
What to do when the business (NLNZ) and the vendor (Ex Libris) can’t agree on an issue
Stakeholders

**NDHA Cross Government Group**

- ACC
- Archives New Zealand
- Crown Law Office
- Customs NZ
- Department of Building & Housing
- Department of Child, Youth & Family Services
- Department of Internal Affairs
- Department of Labour
- Department of the Prime Minister & Cabinet
- Inland Revenue Department
- Land Information New Zealand
- Ministry for Culture & Heritage
- Ministry for the Environment
- Ministry of Agriculture & Forestry
- Ministry of Defence
- Ministry of Economic Development
- Ministry of Education
- Ministry of Justice
- Ministry of Research, Science & Technology
- Ministry of Social Development
- NZ on Air
- Parliamentary Library
- State Services Commission
- Statistics New Zealand
- Te Puni Kōkiri
- The Film Archive
- The Treasury

“anyone who is affected by – or can influence, the decision or outcome”
Gap Analysis

- Analysis of the gap between the Library’s functional requirements for digital preservation and the level of functionality available in the Ex Libris DigiTool product

- 9 categories of gaps
  - Deposit
  - Staging
  - Data model
  - Repository
  - Validation
  - Hardware
  - Permanent
  - Ingest
  - Preservation

72 Gaps
• **G17** - Some of the events should be stored as provenance information.

• At present, 83 separate events are required to be monitored within the system.

• Information relating to some of these events is required to be kept permanently as part of the provenance data for an object going into the future.

**Gap Analysis**

Quite a bit as it turns out ;-)
Gap Analysis
Such as a range of provenance events.

Event 4  – System generates fixity information/values
Event 17 – System deletes a viewing copy
Event 20 – System performs a virus check on a SIP/manifestation
Event 22 – System updates the file fixity Metadata
Event 24 – System validates the file format
Event 38 – Create viewing copies of the files
Event 42 – System creates a manifestation
Event 50 – System deletes a manifestation from archival storage
Event 54 – System stores metadata in the repository
Event 57 – System performs a fixity check on a SIP/manifestation
Event 79 – System runs preservation action on manifestation
### Out of scope gaps

10 gaps were explicitly stated to be out of scope.

<table>
<thead>
<tr>
<th>Hardware</th>
<th>Third party applications</th>
<th>Infrastructure</th>
<th>Repository</th>
</tr>
</thead>
<tbody>
<tr>
<td>NG1 - support for backup</td>
<td>NG5 - hardware support for external components</td>
<td>NG6 - ability to monitor the system</td>
<td>NG7 - support for generating reports on usage and other info</td>
</tr>
<tr>
<td>NG2 - support for disaster recovery</td>
<td>NG8 - support for the Library’s internal systems</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NG3 - support for disaster recovery (off-site backup)</td>
<td>NG9 - support for conversion of current applications into the preservation system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NG4 - any component in the hardware configuration should be duplicated</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Out of scope**

Hardware

NG1 - support for backup
NG2 - support for disaster recovery
NG3 - support for disaster recovery (off-site backup)
NG4 - any component in the hardware configuration should be duplicated

Third party applications

NG5 - hardware support for external components
NG8 - support for the Library’s internal systems
NG9 - support for conversion of current applications into the preservation system
NG10 - support for internal access applications

Infrastructure

NG6 - ability to monitor the system

Repository

NG7 - support for generating reports on usage and other info
Phased Approach

Phase 1 – **Live Production System** 75%
- First delivery of system for acceptance testing
- Testing
- Training
- Phase I Acceptance

NDHA operational for INGEST STORAGE ACCESS

Phase 2 – **Fully Operational** 100%
- Delivery of system for acceptance testing
- Testing
- Training
- Phase 2 Acceptance

NDHA operational for DIGITAL PRESERVATION plus the balance of the functionality

GO LIVE
Work Streams

**Business Change**

**Requirements Management**
FRS
Gap Analysis

Scope / Deliverables Management - Change Control

**Systems Development**
Specify - Code - Unit Test - System Test - Performance Test

DPS R1

**Systems Integration**
CMS - Resource Discovery - Deposit Applications

DPS R2

**Acceptance Testing**
Criteria - Planning - Scenario - Scripting - Test Data - Testing

- July 2004
- Dec 2006
- June 2007
- Sept 2007
- Oct 2008
- Oct 2009
DPS functionality at Day 1

Phase 1 Specifications

From producer management ➔ workflow automation ➔ delivery, audit trails & reporting

- 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
- DPS transformers
- Deposit Application Programme Interface (API)
- Audit & provenance
- Process management
- User management API
- Permanent repository
- Delivery
- Meditor
- Reports
- Back office configuration
Phase 2 delivery

Phase 2 Specifications

Preservation planning and action are in Phase 2 to allow for extended requirements analysis prior to development.

- Format Library
- Risk analysis
- Preservation action
- Enhanced set import/export management for preservation actions
- Maintenance and management functions in Staging NOT permanent
- Enhanced configurability
It’s not just hardware and software

Digital preservation requires interaction with all the organisation’s processes and procedures -

- Business Processes
  - workflows, procedures and policies
- Capacity & Capability
  - resources and skills
- Performance Measures
  - reporting and measuring
- Internal Training
  - system & staff training
- Producer Management
  - service, marketing & training
- Business & Technical Support
  - between departments
- Communication
  - a constant

Organisational readiness

Legislation and strategies are not sufficient

‘No job will be unchanged’
Chief Executive/ National Librarian
October 2007
The NDHA challenge

**Challenge**

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

**Organisational readiness**
Resources, services and infrastructure supporting digital preservation

**Integration with existing systems**

![INDIGO](image)

![WEB CURATOR TOOL](image)

**Migration of digital assets**
80,000 intellectual entities made up of around 280,000 files

**Measuring success**
60 key performance measures
Integration

Integration work stream

It’s not all about the Digital Preservation System

- Deposit applications development
- Existing collection management systems integration
- Browser based content delivery systems development
- Existing resource discovery and delivery systems integration
- Reporting systems
- Common services integration
- Data migration
Integration so far

Milestones

- Staff deposit application
- HTTrack to ARC converter utility
- Archived website migration tools
- OMS data migration tools
- Content aggregator
- Delivery viewers
Internal Submission Application

- Submission Information Package (SIP) Creation Tool (Templates, Hotkey support)

Packages up

- Files (supports complex digital objects)
- Metadata (Structure map creation – METS)
- Digital object structure – multiple representations
- Fixity generation (MD5)
- Links to descriptive record – CMS integration
- Links producer records

Submits SIP to the NDHA
Object Management System (OMS) 2005

Migration 1

- Supporting Legal deposit while NDHA is developed
  - 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
  - Material will be migrated as new content

What to do with all our digital data
Migration 2

Object Management System (OMS) 2005
Data for testing end-to-end DPS functionality

• This will:
  – test initial workflows and process configurations
  – Impose the same metadata constraints (referential integrity, data validation)
  – Impose the same validation checks (fixity verification, virus check, format identification and metadata extraction)
  – Impose the same enrichment tasks (CMS identifier association, access derivative generation)

that will be applied in a live operational setting.

• This should give an indication of the amount of effort required to migrate the rest of the National Library’s digitized content into the NDHA system.
Performance Measures

How do we measure what we’re doing?

From widgets to outcomes

• A move to management information with over 60 key performance indicators including:
  – Key performance indicators
  – Reporting
  – Audit
  – Internal ingest
  – + response actions, i.e. for over/under delivery
Maturity

How ready is our infrastructure for digital preservation?
Business representatives connect the Library to the project

Library

Business Representatives

NDHA Project
Staff Representation on the Project

- Processes
- People
- System
- Information

Training

Sustainable change
Subject Matter Experts (SMEs)

- Ensuring that the NDHA solution and business change take into account Library needs
- Lead analysis as and when required
- Library experts on the NDHA
### Business process

<table>
<thead>
<tr>
<th>Envision</th>
<th>Enable</th>
<th>Enact</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Map current processes✓</td>
<td>• Review gap between current and future processes</td>
<td>• Support local implementation</td>
</tr>
<tr>
<td>• Design NDHA processes</td>
<td>• Contribute to development of high level implementation plan</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Support managers, curators and team leaders in developing local implementation plan</td>
<td></td>
</tr>
</tbody>
</table>
Performance measures

- Envision
  - Propose measurement methodologies
- Enable
  - Support implementation
- Enact

Most Business Representatives
Internal training

Envision
• Advise on and review required adaptations to training materials and exercises for target audience
• Receive training

Enable

Enact
• Deliver training
• Provide NDHA expertise to Department
Producer management

Most Business Representatives

Envision
- Analyse current producers
- Identify future producers
- Develop producer strategies

Enable
- Prepare communications for producers
- Contribute to development of generic producer workflows
- Develop individual producer workflows
- Identify producers for testing
- Adapt training plans and exercises for target audiences
- Receive training

Enact
- Deliver producer training
- Set up individual producer workflows
- Update training programme following evaluation
<table>
<thead>
<tr>
<th>Most Business Representatives</th>
<th>Envision</th>
<th>Enable</th>
<th>Enact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review support requirements</td>
<td>Contribute to development of SLAs</td>
<td>Support implementation</td>
<td></td>
</tr>
<tr>
<td>Review current SLAs</td>
<td>Design client support arrangements</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Develop support plan</td>
<td>Design escalation arrangements</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Design transitional arrangements</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A National strategy
New Zealand’s Digital Strategy: the three Cs
The ‘C’ framework

**connection**

New Zealand will be in the top quarter of the OECD for broadband uptake by 2010.

**content**

To unlock New Zealand’s stock of content and provide all New Zealanders with seamless, easy access to the information that is important to their lives, businesses, and cultural identity.

**confidence**

To provide all New Zealanders with the digital skills and confidence to find and use the information they need; and to ensure that telecommunications and the Internet in New Zealand are reliable and secure.
Five-Element Framework

The power of five

Creating and protecting content
Accessing and discovering content
Sharing and using content
Managing and preserving content
Understanding and awareness of content
The Digital Space

A classification system

Content as:

Formal/Informal

Public/Private

- **Formal**
  - Customer databases, banking, inventories, contracts, tax records, market research, proprietary knowledge, cultural and traditional knowledge.
  - Public directories and records, official information, library catalogues, legislation, archives and heritage collections, open source, scientific and social research.

- **Informal**
  - Correspondence and email, chat and instant messaging, family and holiday snapshots, personal and family memoirs, hobbies and creative pursuits.
  - Blogs and personal webspaces, social networks, forums, discussion boards, wikis, entertainment and news media, tagging, memes.
New Zealand Ideas

Digital New Zealand

- Joined up government
- Public access to legislation
- KRIS – NZ science and research
- Research data sets
- Curriculum resources
- The public record

Digital New Zealand

- Aotearoa People’s Network - community repositories
- Access Radio
- NZ Film Archive
- NZ Sound Archive
- Te Ara – The Online Encyclopedia
- Broadcasting
- Television
Developing a NZ framework

Unifying search and discovery layer

Digital NZ

NDHA
Summary
The ‘C’ framework

**connection**
New Zealand will be in the top quarter of the OECD for broadband uptake by 2010.

**content**
To unlock New Zealand’s stock of content and provide all New Zealanders with seamless, easy access to the information that is important to their lives, businesses, and cultural identity.

**confidence**
To provide all New Zealanders with the digital skills and confidence to find and use the information they need; and to ensure that telecommunications and the Internet in New Zealand are reliable and secure.

**continuity**
Digital preservation is the glue which will make this happen and where sustainability of our work in the digital domain occurs.
Collecting, preserving and making accessible New Zealand’s published digital heritage
Challenges Ahead

Digital preservation must be a national issue

Does digital preservation require an international response?

- Digital preservation is at the heart of our business in 5 years
- Citizen’s created content impacting on our collection, description and preservation processes
- Content (ie digital preservation) systems are our core operational systems, not the catalogue
- Digital preservation as a component of a national knowledge infrastructure
- Quality assurance of products and tools
- Professional services market
- Agreed lexicon describing what we mean by digital preservation and what we want from digital preservation systems

Does digital preservation require an international response?
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.
What is a national library for?

“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.”

Thank You

Steve Knight
Associate Director National Digital Library
National Library of New Zealand