Digitálna Knížnica 2009
Jasná, March 16 2009

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

Call to action
Wait or act now

- Legislative mandate and strategic context

- Development of digital preservation expertise

- International library community
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.”
'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.'
Components of digital preservation:

• Storage
• Psychology of digital preservation
• Digital preservation processes/strategies
• This is only the beginning

• It’s important to know what you want to achieve
  • what are your requirements when you talk about digital preservation
  • make sure you understand the domain and make the decisions for your institution
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.
Collaboration

The NDHA Programme will be successful and delivered in a timely and cost effective manner.

Partnership

Design & Build

Sun Centre of Excellence

Ex Libris
FROM LIBRARY SYSTEMS TO INFORMATION SERVICES

Sun Microsystems
OAIS Reference Model
Development decision

Goal
Collection, preservation and access in perpetuity

- 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
The Solution

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

NDHA is the NLNZ implementation of DPS
- wider functionality and business change are required for practical digital preservation within any given institution
**Work Streams**

### Business Change


### Requirements Management

FRS

Gap Analysis

Scope / Deliverables Management - Change Control

---

**Systems Development**

Specify - Code - Unit Test - System Test - Performance Test

**Systems Integration**

CMS - Resource Discovery - Deposit Applications

**Acceptance Testing**

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

---

**DPS R1**

**DPS R2**

---

**Dates:**
- July 2004
- Dec 2006
- June 2007
- Sept 2007
- Oct 2008
- Oct 2009

---

**Special Notes:**
- June 2007
- July 2007
- Sept 2007
- Oct 2007
- Oct 2008
- Oct 2009

---
Business Change

Workstream

‘No job will be unchanged’

Chief Executive/National Librarian

October 2007

- 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
DPS Functionality at Day 1

Phase 1
Specs

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
<table>
<thead>
<tr>
<th>Specification</th>
<th>Covers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Producers</td>
<td>Establish and maintain accounts for NDHA producers, with information that is pertinent for preservation of and reporting about digital content from each producer.</td>
</tr>
<tr>
<td>Deposit 1</td>
<td>Registration and management of individual depositor accounts and associations linking each with a Producer account.</td>
</tr>
<tr>
<td>Deposit 2</td>
<td>Establish and make use of (by depositors) deposit flows and forms that govern the deposit of digital content, covering web deposit, staff-mediated deposits, internal bulk deposits and external bulk deposits.</td>
</tr>
<tr>
<td>Validation Stack</td>
<td>Automated processes applied at time of ingest. Includes format identification, extraction of technical metadata, fixity and virus check services (integrity checking).</td>
</tr>
<tr>
<td>Specification</td>
<td>Covers</td>
</tr>
<tr>
<td>------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>IE Data Model</td>
<td>The agreed digital domain and data models that provide the essential foundation for the digital repository and preservation system.</td>
</tr>
<tr>
<td>SIP Submission</td>
<td>Collectively, these specifications cover the system processes and workflows that are required for ingesting new content. Covers automated ingest, ‘enrichment’ routines and staff intervention for addressing technical issues affecting deposited content, selection / arrangement (approval) needed prior to storage in the archive’s permanent repository.</td>
</tr>
<tr>
<td>SIP Processing</td>
<td></td>
</tr>
<tr>
<td>Registration</td>
<td></td>
</tr>
<tr>
<td>Technical Analyst</td>
<td></td>
</tr>
<tr>
<td>Assessor</td>
<td></td>
</tr>
<tr>
<td>Approver</td>
<td></td>
</tr>
<tr>
<td>Arranger</td>
<td></td>
</tr>
</tbody>
</table>
### Specifications

<table>
<thead>
<tr>
<th>Specification</th>
<th>Covers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Set Management</strong></td>
<td>Staff intervention to carry out maintenance activities on discrete instances of content held in the archive. Set management also supports preservation planning and actions that will be delivered in the Phase 2 system.</td>
</tr>
<tr>
<td><strong>Directory Structure Transformers</strong></td>
<td>Includes pre-transformers for converting non-standard content structures into standard content structures and transformers for converting standard content structures into SIP METS format.</td>
</tr>
<tr>
<td><strong>GO API</strong></td>
<td>Enables bypassing of the Deposit UI/Deposit Client for submission of batches of METS in a predefined format directly into the Deposit Server.</td>
</tr>
<tr>
<td><strong>Audit Trail</strong></td>
<td>Full audit records and attachment of provenance events affecting each digital object taken into the archive’s permanent repository.</td>
</tr>
<tr>
<td>Specification</td>
<td>Covers</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Workflow / Process Automation</td>
<td>Management and configuration of system workflows and automated processes.</td>
</tr>
<tr>
<td>Staff Management</td>
<td>Definition of roles available within the NDHA system and assignment of roles to Library staff who need to access and use the NDHA system for permitted actions.</td>
</tr>
<tr>
<td>User Management</td>
<td>Definition of roles available within the NDHA system and assignment of roles to external Users who need to access and use the NDHA system for permitted actions.</td>
</tr>
<tr>
<td>Permanent Repository</td>
<td>Final storage location of objects in the system. Provides services such as Delivery and Publishing and informs preservation risk analysis and preservation actions.</td>
</tr>
</tbody>
</table>
## Specifications

<table>
<thead>
<tr>
<th>Specification</th>
<th>Covers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delivery</td>
<td>Determining the exact/proximate line between the DPS and institutional delivery/presentation softwares and the level of integration expected.</td>
</tr>
<tr>
<td>Meditor</td>
<td>The suite of object and metadata editing capabilities (interface screens and flows) and the point of intersection with other DPS tasks, eg Set Management which identifies where Meditor functions are invoked.</td>
</tr>
<tr>
<td>Reports</td>
<td>Management, operational and statistical reporting, using views of the object data model.</td>
</tr>
</tbody>
</table>
Integration

Workstream

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
What is INDIGO

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
Challenges going forward

• Language - Requirements
• The data deluge
• Products, tools and services
• Quality assurance and confidence
• Drivers towards standards/ best practice
• Challenges ahead
More data deluge

**BBC**
Petabytes per week

**CERN LHC – black holes (mini or otherwise)**
How much data?

Content complexity
- Kam Woods – CDs
- Alex Ball – CAD (Engineering)
- Mark Guttenbrunner (gaming)
Digital preservation is at the heart of our business

- Agreed lexicon describing what we mean by digital preservation and what we want from digital preservation systems
- Capability/capacity to respond to technological change and innovation
- Quality assurance of products and tools – enterprise class software
- Professional services market
- Defining, resourcing and pursuing the research agenda
- Digital preservation as a component of a national knowledge infrastructure
How should we Describe Digital Preservation?

Components of digital preservation:

• Storage
• Psychology of digital preservation
• Digital preservation processes/strategies
• This is only the beginning

• It’s important to know what you want to achieve
  • what are your requirements when you talk about digital preservation
  • make sure you understand the domain and make the decisions for your institution
End result – a fully functioning National Digital Heritage Archive as at 1 November 2008

With

Phase 2 due in 2009
Why we Think this Matters

Thank you

“A people without knowledge of its past is … like a tree without roots.”

Marcus Garvey

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