Safe Havens in a Choppy Sea: Digital Object Management Workflows at the National Library of Australia

Gerard Clifton
Manager, Digital and Audio Preservation Resources
National Library of Australia
**Seascape: NLA Digital Collections**

- Digitised materials
  - Images
  - Sound recordings
- Online materials (PANDORA)
  - Online publications, Web sites
- Published offline materials
  - Documents, maps & multimedia
  - Floppy disc, CD, DVD
- Manuscript materials in digital form
  - Digital manuscripts, papers, email
Presentation Overview

- NLA Digital Services Architecture
- Persistent identifiers
- Storage
- Collection and management workflows
  - Digital Collections Manager (Digitisation)
  - PANDAS (Web archiving)
  - Future work
Digital Services Architecture

Principles:

• Support the entire activity cycle
  – collection, storage, management, preservation, discovery, access and delivery
• Integrate access to print & digital materials
• Support hierarchical model
• Provide contextual data, navigation, delivery
• Persistent citation and access

Cathro & Boston (2003) :

Digital Services Architecture

- **Discovery**
  - ILMS OPAC
  - Metadata Repository & Search System

- **Persistent Access**
  - Resolver Service

- **Delivery**
  - Delivery Systems
    - Digital Collections Manager (DCM) & Digital Archiving System (PANDAS)
    - Digital Object Storage System (DOSS)

- **Collection & Management**

- **Storage**
Persistent Identifiers

- Support for versions


---

- Role codes (examples)

<table>
<thead>
<tr>
<th>Role Code</th>
<th>Description</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>o</td>
<td>Examination copy</td>
<td>ms51-13-1296</td>
</tr>
<tr>
<td>m</td>
<td>View copy</td>
<td>nla.ms</td>
</tr>
<tr>
<td>c</td>
<td>Thumbnail</td>
<td>s2</td>
</tr>
<tr>
<td>t</td>
<td>Related metadata</td>
<td>t</td>
</tr>
<tr>
<td>dm</td>
<td>Related metadata</td>
<td>v1</td>
</tr>
</tbody>
</table>
Delivery System


More Information

Digital Services Architecture

- Discovery
  - ILMS OPAC
  - Metadata Repository & Search System
- Persistent Access
- Delivery
  - Resolver Service
  - Delivery Systems
- Collection & Management
  - Digital Collections Manager (DCM) & Digital Archiving System (PANDAS)
- Storage
  - Digital Object Storage System (DOSS)
Digital Object Storage

DOSS System Architecture

- Oracle Server E250
- Apache Web Server E250
- DOMS/ASM Server E450
- Brocade Switches
- Navisphere Win 2000
- Clarion 4700 Disk Array
- Derivatives (delivery)
- STK L700 Tape Library
- Masters
- ACSLS Server Ultra5S
- NLA LAN

Capacities: Disc 14 TB, Tape 60 TB

Digital Collections Workflows

- Digitised images (pictures, maps, manuscripts, music)
- Audio materials
- Sound Preservation
- Digital Audio Workstations
- Digital Collections Manager (DCM)
- Online materials
- Digital Archiving Section
- PANDAS
- Digital Object Storage System (DOSS)
Digital Collections Manager

- Supports digitisation workflows and object management
  - upload preservation masters
  - create delivery derivatives
  - manage file storage
  - store descriptive, technical, process metadata
  - find & browse works, copies, jobs, projects
  - upload, abort, download
Digital Collections Manager

Work hierarchies

Copy & file characteristics & relationships

Process histories

Digitisation Workflow

1. Create Work & Copy details
2. Request Job
3. Create Digital Content
4. Upload to storage
5. Acquit Job

- Work details (& hierarchies) can be:
  - Imported from catalogue
  - Imported from finding aids in EAD
  - Created within DCM

- Add Copy details for each Work
  - Characteristics of each copy

- Assign persistent identifiers
  (based on pattern for collection)
Digitisation Workflow

1. Create Work & Copy details
2. Request Job
3. Create Digital Content
4. Upload to storage
5. Acquit Job

- Search DCM for items
- Add selected items to ‘Cart’
- Create Job
Digitisation Workflow

1. Create Work & Copy details
2. Request Job
3. Create Digital Content
4. Upload to storage
5. Acquit Job

- Born-digital Originals
  - Images (TIFF)
  - Digital audio recordings (WAV or BWF)

- Digitised content (from physical originals)
  - Digitised from selected copy
  - Professional equipment & standards
    - Images: scanners, cameras (TIFF)
    - Audio: tape and DAT replay, A/D converters, sound cards (WAV)

- QUADRIGA digital audio workstations convert WAV to BWF
Digitisation Workflow

1. Create Work & Copy details
2. Request Job
3. Create Digital Content
4. Upload to storage
5. Acquit Job

- Special step for audio content
- ‘Preprocess’
  - creates XML metadata file for conversion of WAV to BWF by QUADRIGA workstation
Incorporation of metadata into BWF

Digital Collections Manager (DCM)

Quadriga DAW
Create Digital Content

Incorporation of metadata into BWF

Digital Collections Manager (DCM)

XML A

Quadriga DAW
Incorporation of metadata into BWF
3 Create Digital Content

Incorporation of metadata into BWF

- Digital Collections Manager (DCM)
- XML A
- Subset of A
- Quadriga DAW
- BWF
Create Digital Content

Incorporation of metadata into BWF

Diagram showing the relationship between Digital Collections Manager (DCM), XML A, XML B, Quadriga DAW, BWF, and a subset of XML A.
Create Digital Content

Incorporation of metadata into BWF
Create Digital Content

Incorporation of metadata into BWF

Digital Collections Manager (DCM)

Upload
BWF

Subset of B

XML A

Subset of A

XML B

Quadriga DAW
Digitisation Workflow

- Content submitted to uploader
  - Check for required files
  - Output copy records created
  - Create version numbers
  - Checksum generated (MD5)
  - Extract metadata
  - Generate delivery copies (& records)
  - Move files to DOSS storage
**Digitisation Workflow**

1. Create Work & Copy details
2. Request Job
3. Create Digital Content
4. Upload to storage
5. Acquit Job

- Metadata recorded
  - Descriptive & Structural metadata
    - From Work & Copy records
    - Subunit details
    - Parent-child relationships
  - Process metadata
    - Actions (Replicate, Derive, Ingest)
    - Input and output copies
    - Tools
  - Technical metadata
    - File system metadata
    - Format-specific metadata
Digitisation Workflow

1. Create Work & Copy details
2. Request Job
3. Create Digital Content
4. Upload to storage
5. Acquit Job

- Jobs acquitted
- Copies made available to system
- (Access restrictions remain in force)
## Examples

### Image – Master copy detail

<table>
<thead>
<tr>
<th>Description</th>
<th>History</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find a work</td>
<td>Result set</td>
</tr>
</tbody>
</table>

Displaying copy details for nla.pic-an24283499-m-v1:

- **Copy ID:** nla.pic-an24283499-m-v1
- **Work ID:** nla.pic-an24283499
- **File name:** nla.pic-an000024283499-m.tif
- **Access:** Unrestricted
- **File format:** image/tiff (TIFF file)
- **File size:** 106,144,586 bytes (101.2 MB)
- **Encoding:** MAC
- **Compression:** None
- **Checksum:** 112dDc2d79d8d49e6663529666e64e8 (MD5)
- **Resolution:** 300 x 300 ppi
- **Bit depth:** 24 bits
- **Colour space:** RGB
- **Orientation:** Portrait
- **Dimensions:** 5,542 x 6,384 pixels
- **Date created:** 3 Feb 2003
- **Source copy:** nla.pic-an24283499-o-v1
- **Agent:** Imaging Services (job no. 5331)

Co-master available
## Audio – Master copy detail

<table>
<thead>
<tr>
<th>Description</th>
<th>Versions</th>
<th>History</th>
</tr>
</thead>
<tbody>
<tr>
<td>Displaying copy details for nla.oh-4599-0047-0001-m-v2.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CopyId: nla.oh-4599-0047-0001-m-v2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WorkId: nla.oh-4599-0047-0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FileName: nla.oh-004599-0047-0001-m.wav</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access: Unrestricted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Track: 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SoundField: Stereo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration: 00:53:39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration Type: Input</td>
<td></td>
<td></td>
</tr>
<tr>
<td>File Format: audio/x-wav (Microsoft RIFF)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>File Size: 610,469,204 bytes (599.3 MB)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compression: None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Checksum: e9810803b4309cdd42ca83b6a767115 (MD5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bit Depth: 16 bits</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sampling Rate: 48000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date Created: 22 Aug 2005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source Copy: nla.oh-4599-0047-0001-v1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agent: SFATS (job no. 27013)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Material type: Sound; Copy Type: Digital; Copy Role: Master; Carrier: Online; Version: 2; Created: 22 Aug 2005 (DCM); Last updated: 23 Aug 2005 (gilliton)
Examples

Audio – Master process history

<table>
<thead>
<tr>
<th>Process</th>
<th>Input copy</th>
<th>Output copy</th>
<th>Status</th>
<th>Date assigned</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicate</td>
<td>rla:oh-4599-0047-0001-m-v1</td>
<td>rla:oh-4599-0047-0001-m-v2</td>
<td>Acquired</td>
<td>22 Aug 2005</td>
</tr>
<tr>
<td>Ingest</td>
<td>rla:oh-4599-0047-0001-m-v2</td>
<td></td>
<td>Acquired</td>
<td>22 Aug 2005</td>
</tr>
<tr>
<td>Derive</td>
<td>rla:oh-4599-0047-0001-m-v2</td>
<td>rla:oh-4599-0047-0001-rb4-v1</td>
<td>Acquired</td>
<td>22 Aug 2005</td>
</tr>
<tr>
<td>Derive</td>
<td>rla:oh-4599-0047-0001-m-v2</td>
<td>rla:oh-4599-0047-0001-ra1-v1</td>
<td>Acquired</td>
<td>22 Aug 2005</td>
</tr>
<tr>
<td>Derive</td>
<td>rla:oh-4599-0047-0001-m-v2</td>
<td>rla:oh-4599-0047-0001-rm-v1</td>
<td>Acquired</td>
<td>22 Aug 2005</td>
</tr>
<tr>
<td>Derive</td>
<td>rla:oh-4599-0047-0001-m-v2</td>
<td>rla:oh-4599-0047-0001-rb2-v1</td>
<td>Acquired</td>
<td>22 Aug 2005</td>
</tr>
<tr>
<td>Derive</td>
<td>rla:oh-4599-0047-0001-m-v2</td>
<td>rla:oh-4599-0047-0001-rb3-v1</td>
<td>Acquired</td>
<td>22 Aug 2005</td>
</tr>
<tr>
<td>Derive</td>
<td>rla:oh-4599-0047-0001-m-v2</td>
<td>rla:oh-4599-0047-0001-sa1-v1</td>
<td>Acquired</td>
<td>22 Aug 2005</td>
</tr>
<tr>
<td>Derive</td>
<td>rla:oh-4599-0047-0001-m-v2</td>
<td>rla:oh-4599-0047-0001-sb2-v1</td>
<td>Acquired</td>
<td>22 Aug 2005</td>
</tr>
<tr>
<td>Derive</td>
<td>rla:oh-4599-0047-0001-m-v2</td>
<td>rla:oh-4599-0047-0001-sb1-v1</td>
<td>Acquired</td>
<td>22 Aug 2005</td>
</tr>
<tr>
<td>Derive</td>
<td>rla:oh-4599-0047-0001-m-v2</td>
<td>rla:oh-4599-0047-0001-rt1-v1</td>
<td>Acquired</td>
<td>22 Aug 2005</td>
</tr>
</tbody>
</table>
Examples

Audio – Master process detail

Displaying process details for the nominated process:

- **Job number:** 27013
- **Process type:** Replicate
- **Process role:** Primary
- **Input copy:** ria.ob.4588-0047-0001-o-v1
- **Output copy:** ria.ob.4588-0047-0001-m-v2
- **Status:** Acquired
- **Date assigned:** 22 Aug 2005
- **Signal source:** DAT-Inspector
Web archiving: PANDAS

PANDAS
(PANDORA Digital Archiving System)

PANDORA Archive
Web Archiving: PANDAS

- Workflow, gathering and management system for Web materials
- Uses HTTrack as gatherer
- PI pattern – reflects gathered URLs:

```
<Collection>  <Title PI>  <Instance Date>  <URL>
```

- nla.arc
- 45920
- 20050222
Web Archiving Workflow

Select

Register

Gain Permissions

Gather
- schedule
- filters

HTTrack crawl

Process
- QA check
- QA fix

Initial Capture

QA Copy

Archive
- Preservation Master (TAR)
- Archive Master (TAR)
- Display copies

Restrict

Catalogue

Set Display
Digital Object Management

• Administration
  – Management of works, copies, relationships, metadata

• Data Management
  – Redundant storage and backup
  – Refreshment cycles
  – Restrictions on access
  – User authentication

• Delivery
  – Persistent citation and access
  – Online delivery
Future enhancements

- **Digital Collections Manager**
  - Additional formats (e.g. PDF)
  - Rights management

- **PANDAS**
  - Batch loading
  - IA Heritrix crawler and ARC format
**Future enhancements**

- **Both**
  - Review repository layer options (e.g. FEDORA)
  - Enhance Preservation Planning and management
- **With APSR:**
  - Enhance support for preservation metadata (PREMIS)
  - Investigate automated monitoring of format obsolescence (PANIC)
- **Support for preservation actions**
Conclusion

• Review our position
• Plot our course
• Plan for weather
• Watch the horizon