

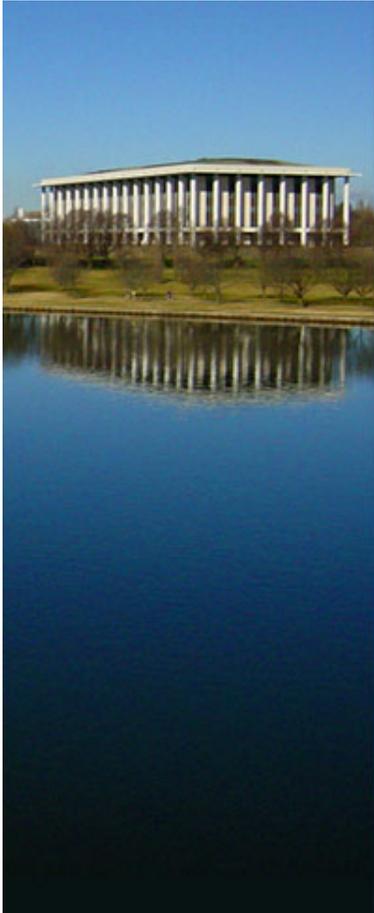


# *Safe Havens in a Choppy Sea:* Digital Object Management Workflows at the National Library of Australia

International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



Gerard Clifton  
Manager, Digital and Audio Preservation Resources  
National Library of Australia



# *Seascope: 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

International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



# *Presentation Overview*

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

International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



# *Digital Services Architecture*

## Principles:

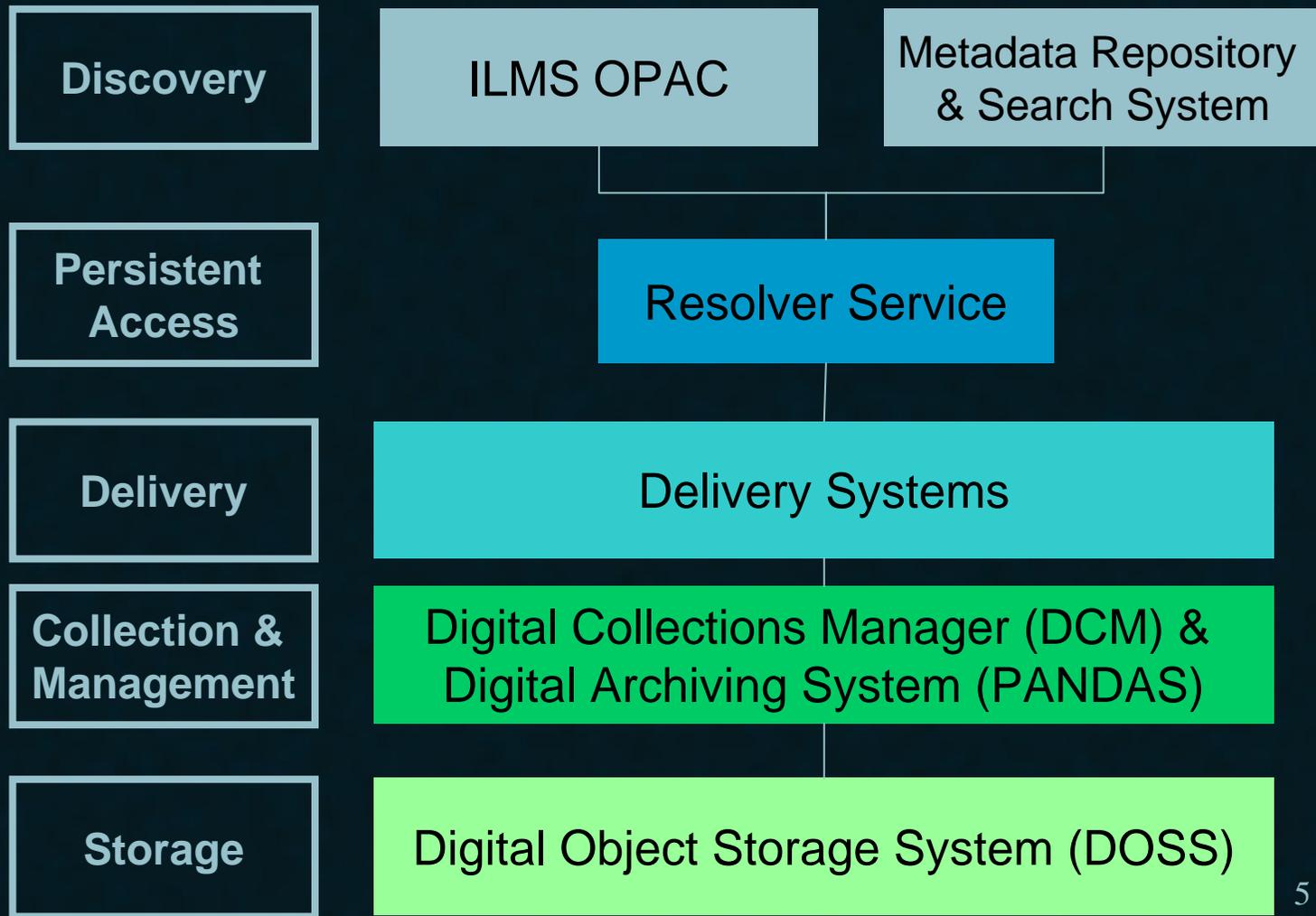
- Support the entire activity cycle
  - collection, storage, management, preservation, discovery, access and delivery
- Integrate access to print & digital materials
- Support hierarchical model
- Provide contextual data, navigation, delivery
- Persistent citation and access

International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany





# Digital Services Architecture

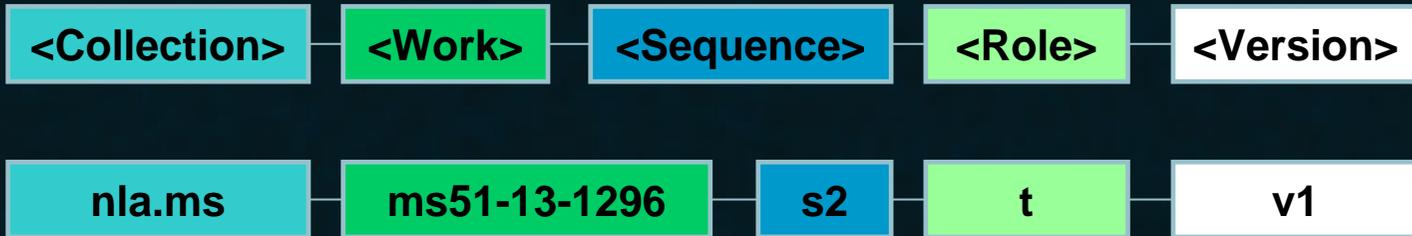


International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany





# Persistent Identifiers



- Role codes (examples)

Original	o	Examination copy	e
Master	m	View copy	v
Co-master	c	Thumbnail	t
Derivative master	dm	Related metadata	rm

- Support for versions

International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany

# Delivery System



Digital Collections - Maps - Coode, John, Sir. Melbourne Harbour Trust [cartographic material] - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://www.nla.gov.au/apps/cdview?pi=nla.map-nk1436> Go Links >>

**Digital Collections**  
Maps

NATIONAL LIBRARY OF AUSTRALIA  
[copyright](#) | [order](#) | [contact us](#) | [help](#)

Coode, John, Sir.  
**Melbourne Harbour Trust [cartographic material] : works of improvement recommended by Sir John Coode in his report of 17th. Feb. 1879.**  
1879. MAP NK 1436.  
[More information](#)

ZOOM IN PRINT Select map GO

[Search](#) > Maps



[Search](#) > Maps Cite page: <http://nla.gov.au/nla.map-nk1436>

ZOOM IN PRINT Select map GO

Local intranet

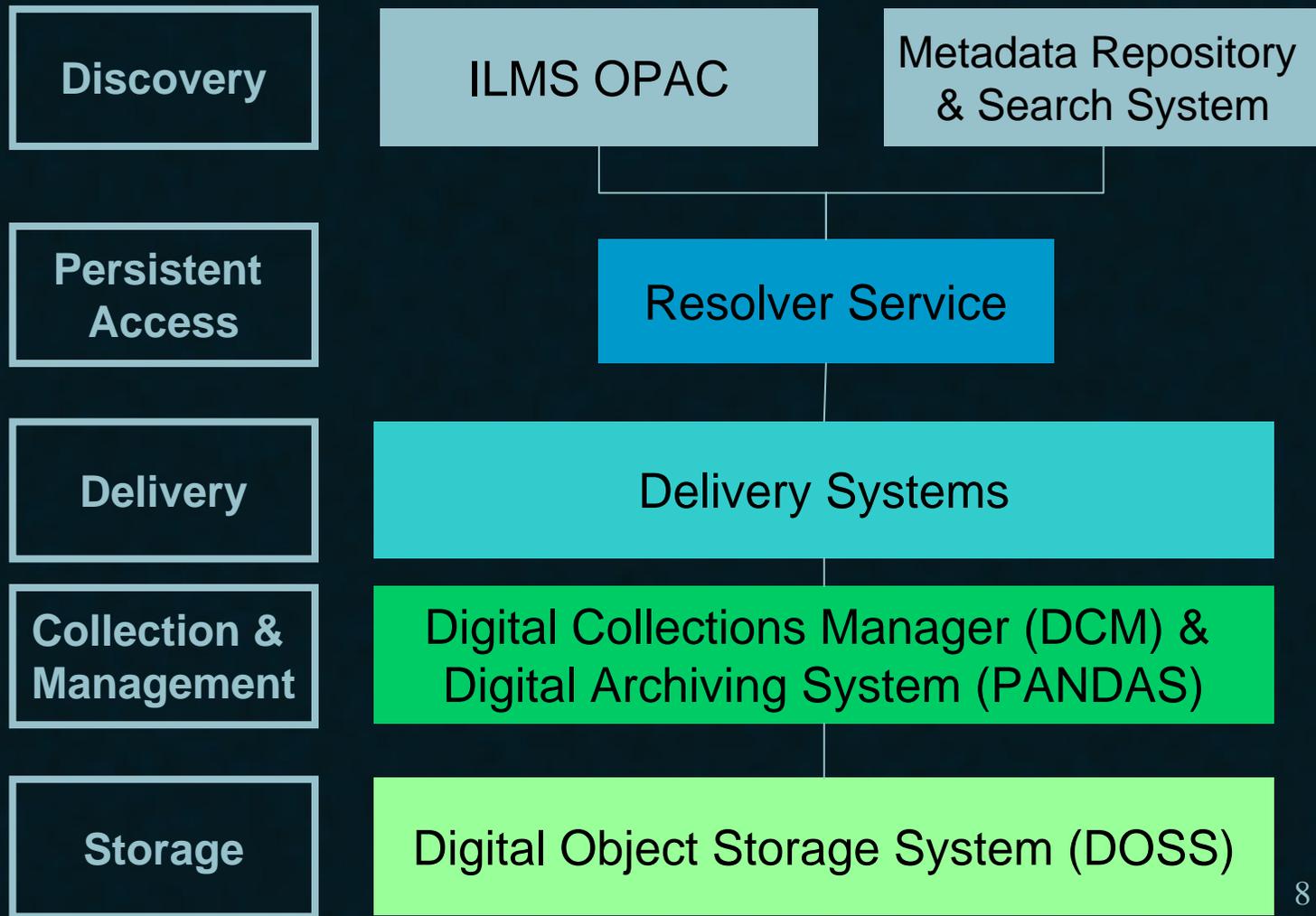
International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



Try this page live at : <http://nla.gov.au/nla.map-nk1436>



# Digital Services Architecture

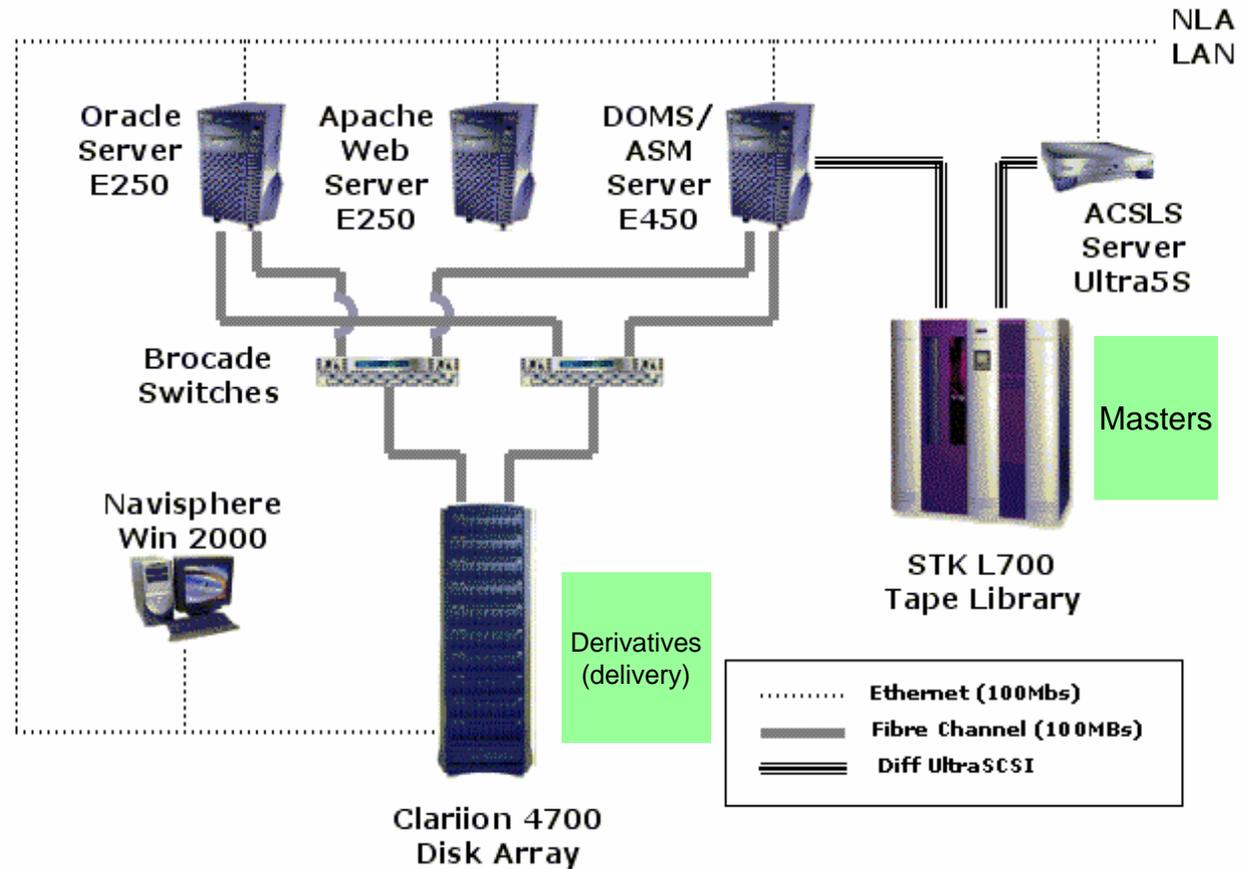


International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



# Digital Object Storage

## DOSS System Architecture



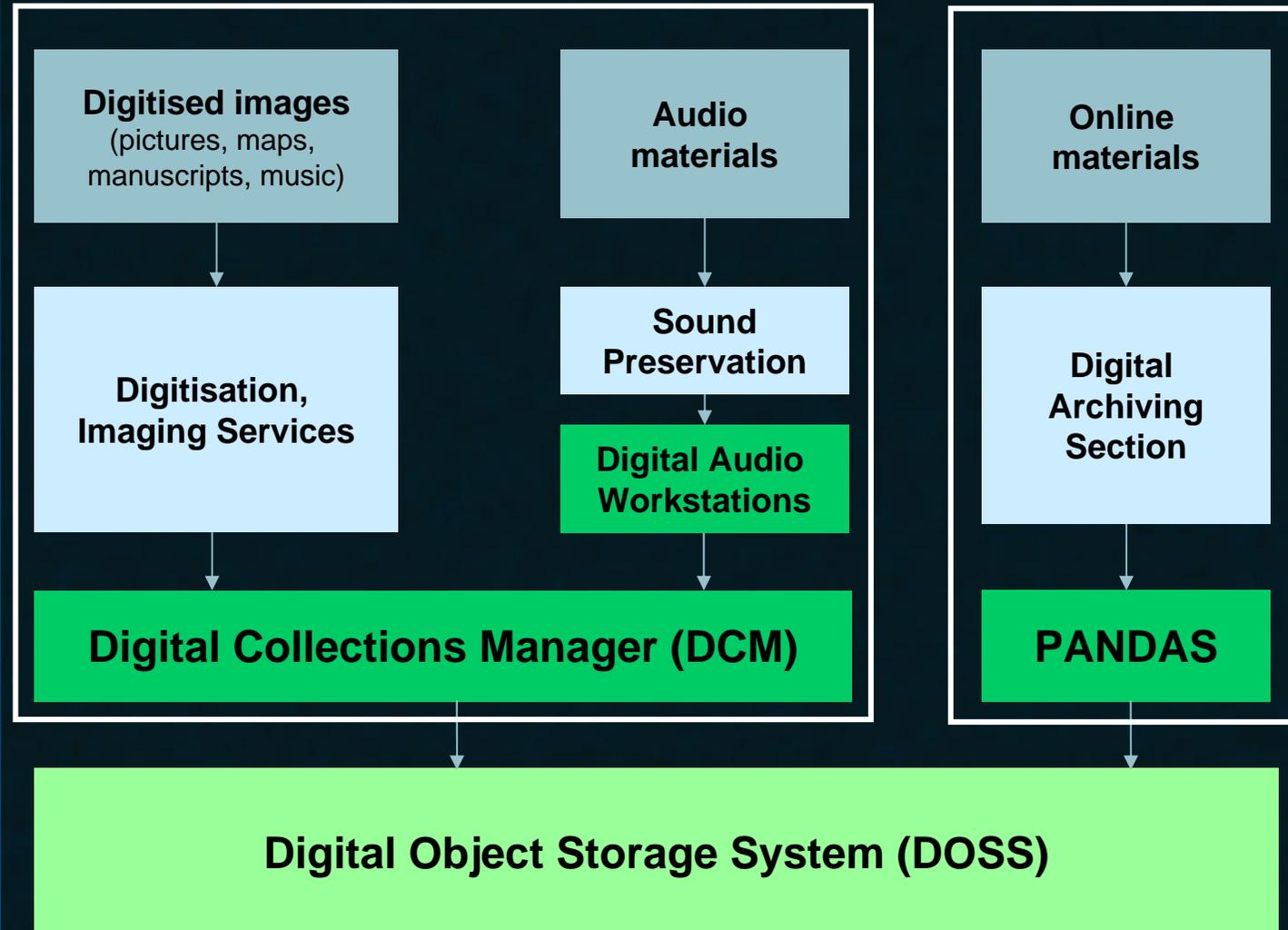
International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



Capacities: Disc 14 TB Tape 60 TB

Further information : <http://www.nla.gov.au/dsp/doss/doss.doc>

# Digital Collections Workflows



International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



# *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

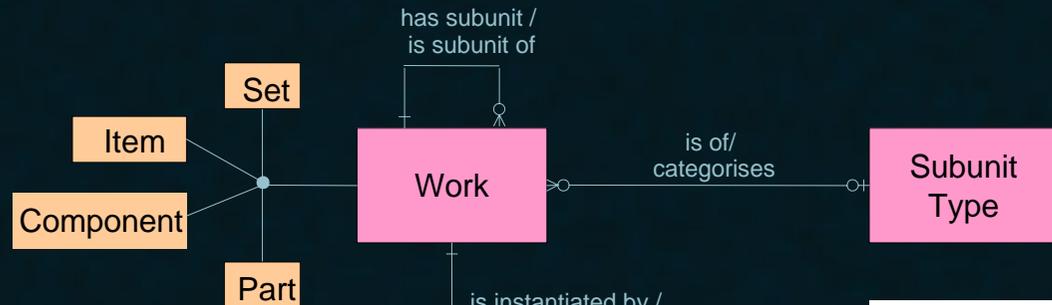
International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



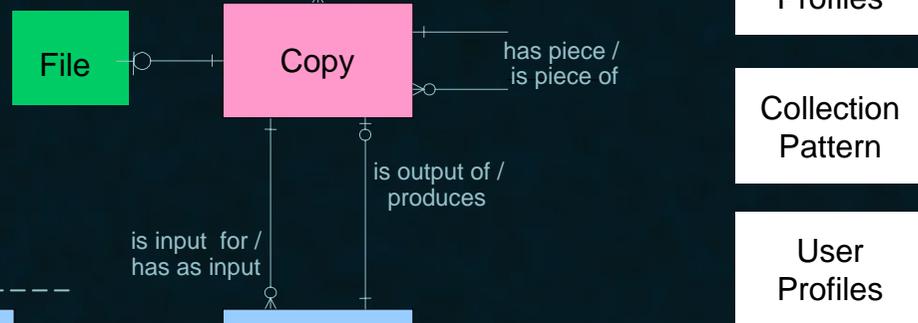


# Digital Collections Manager

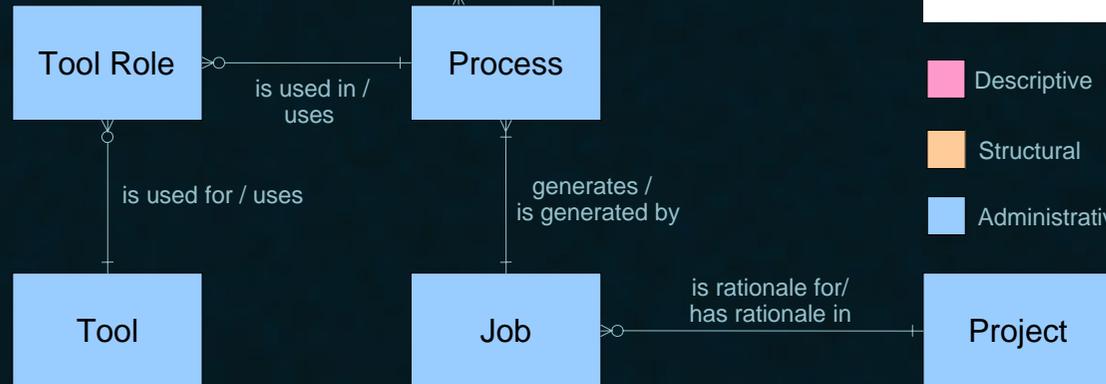
## Work hierarchies



## Copy & file characteristics & relationships



## Process histories



- Descriptive
- Structural
- Administrative

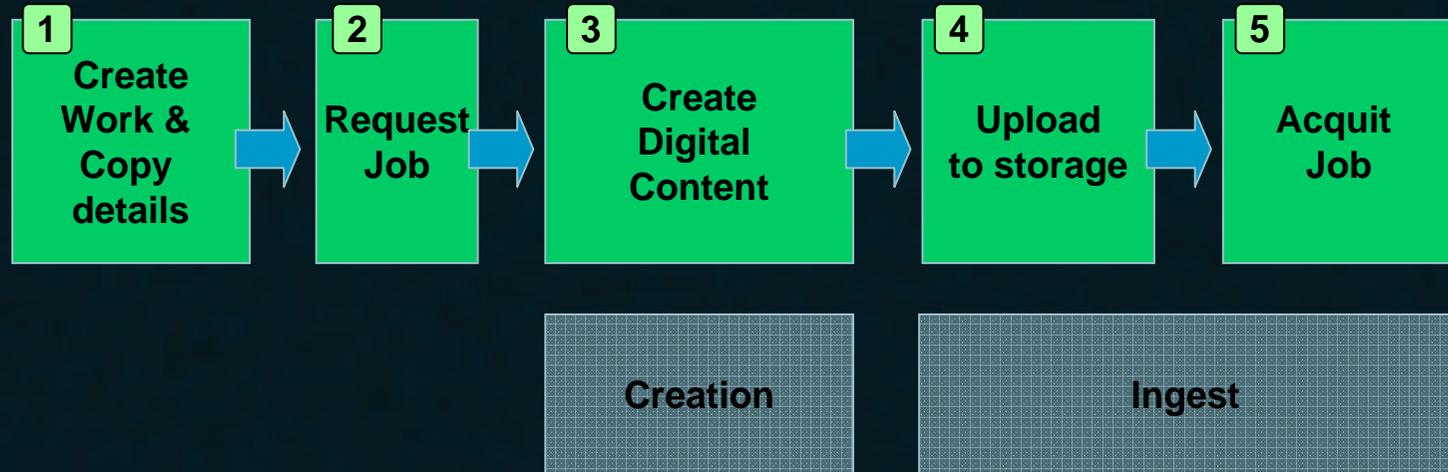
International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



Functional specifications available from:  
<http://www.nla.gov.au/dsp/doms/dcm.html>

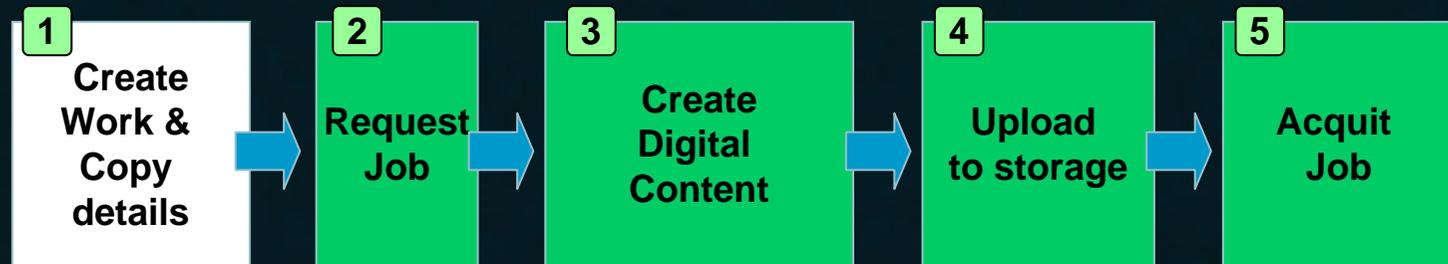


# Digitisation Workflow



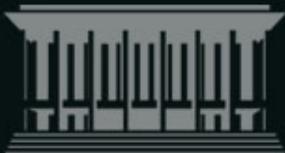
International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany

# Digitisation Workflow



- 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)

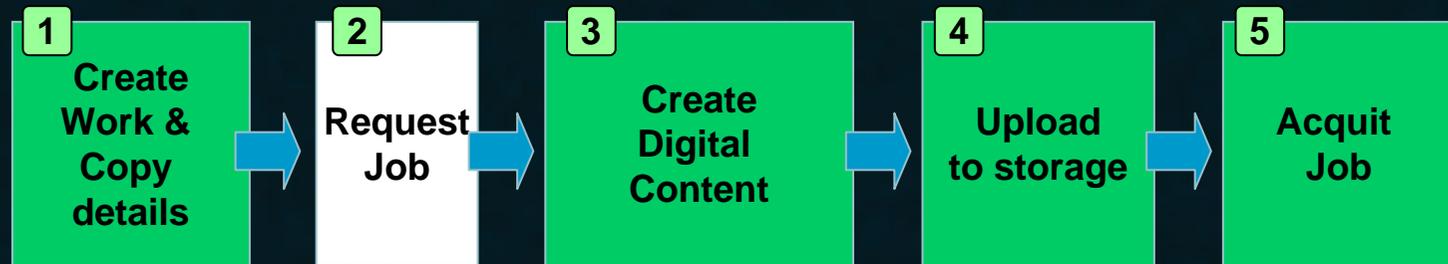
International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



NATIONAL  
LIBRARY OF AUSTRALIA



# Digitisation Workflow



- Search DCM for items
- Add selected items to 'Cart' 
- Create Job

International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



# Digitisation Workflow



- 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

International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



# Digitisation Workflow



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

International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



### 3 *Create Digital Content*

Incorporation of metadata into BWF

Digital Collections  
Manager (DCM)



Quadriga DAW

International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany

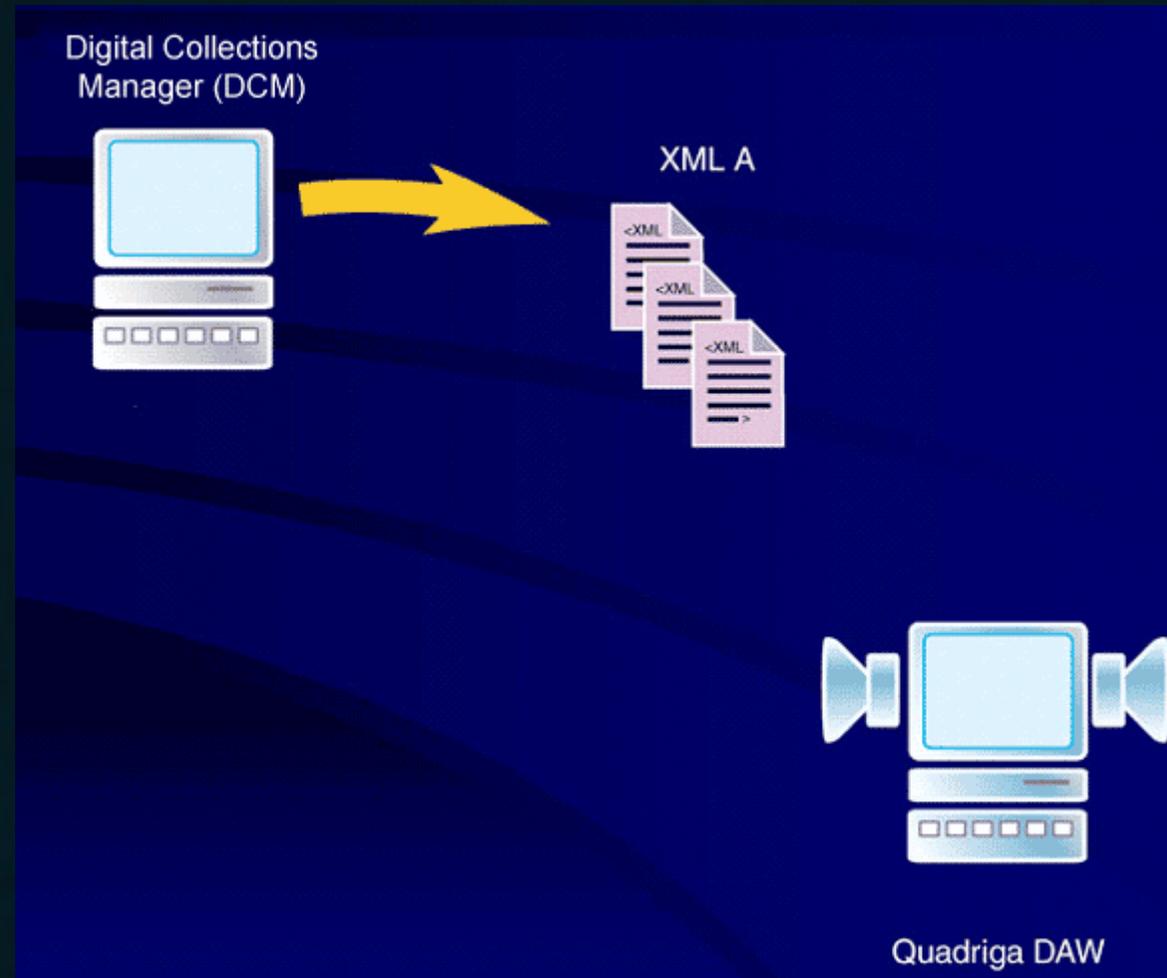


NATIONAL  
LIBRARY OF AUSTRALIA



### 3 *Create Digital Content*

Incorporation of metadata into BWF

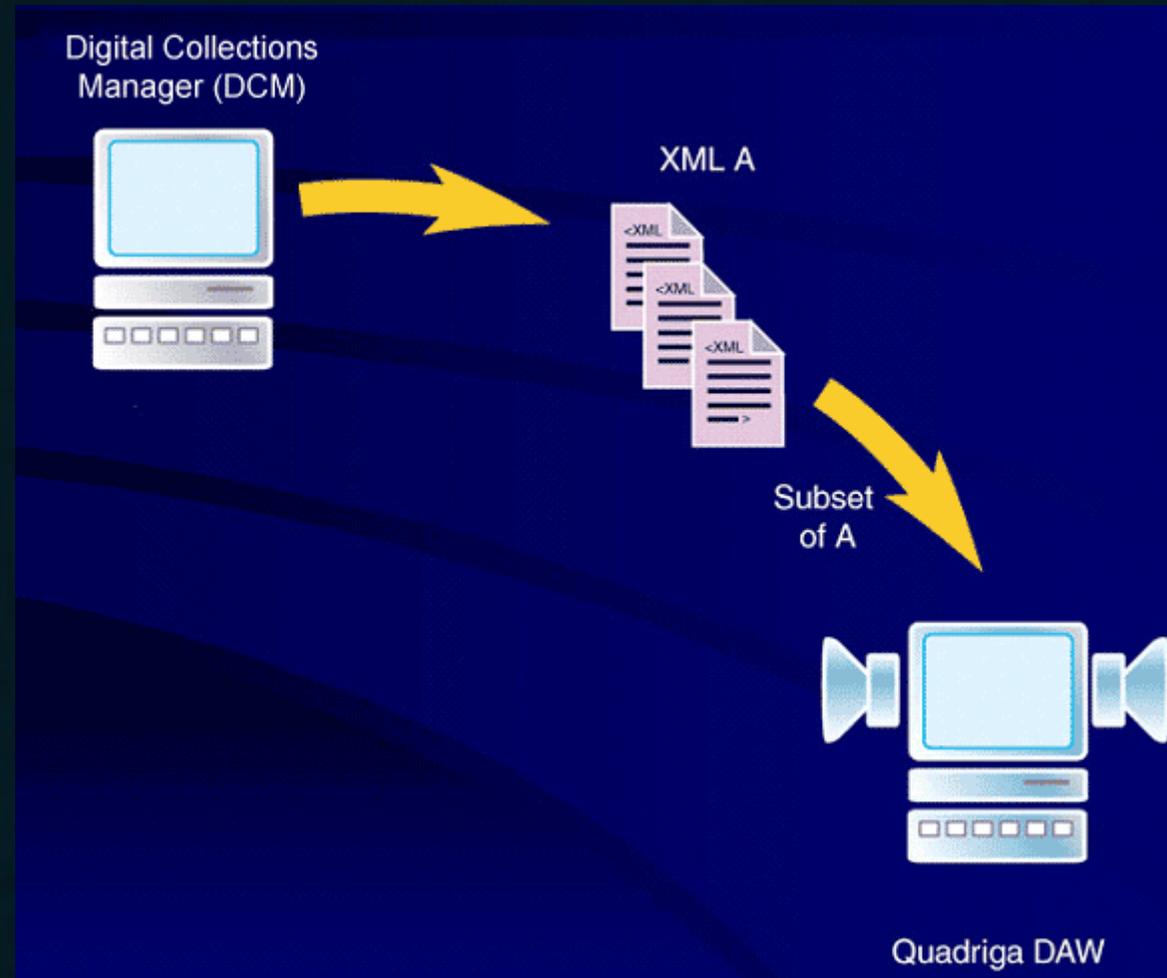


International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



### 3 *Create Digital Content*

Incorporation of metadata into BWF

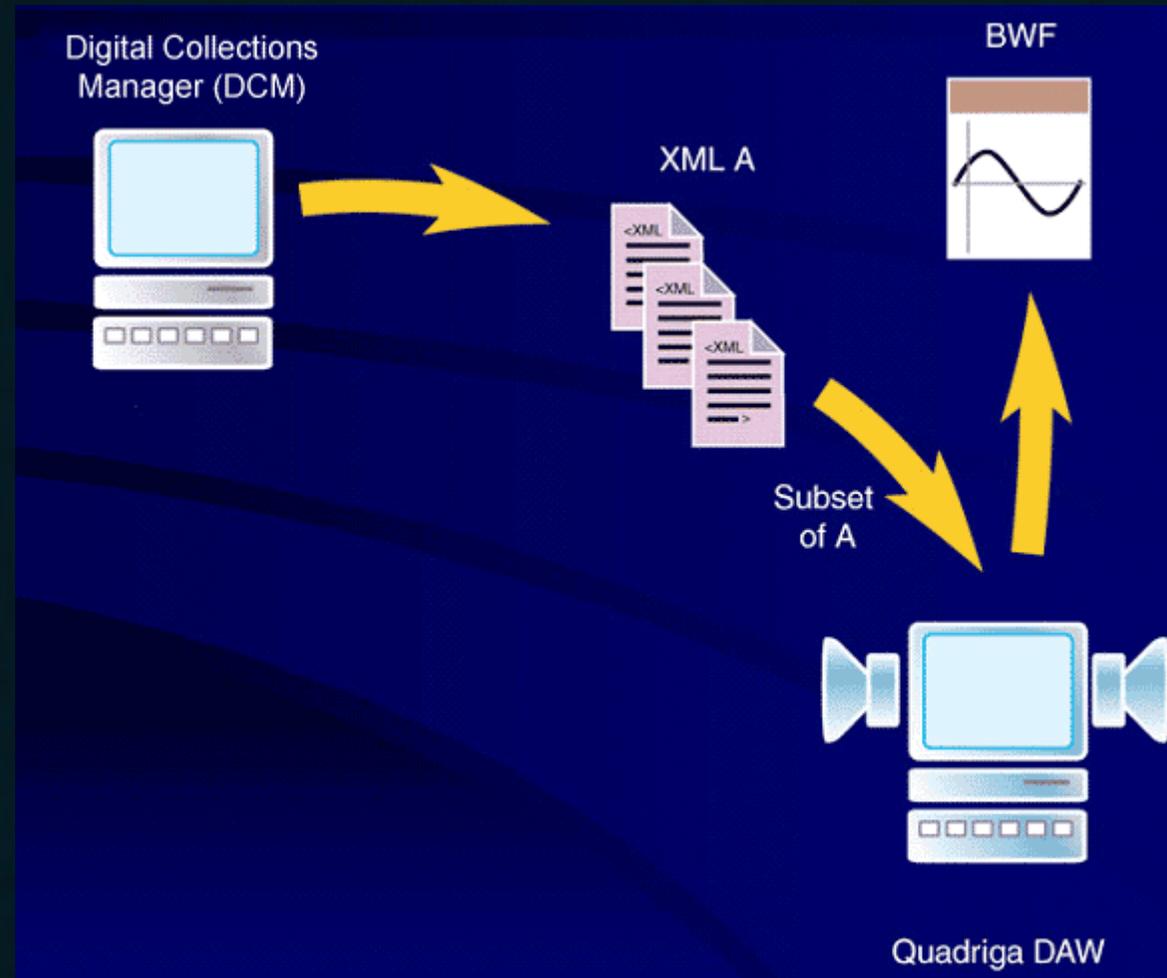


International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



### 3 *Create Digital Content*

Incorporation of metadata into BWF

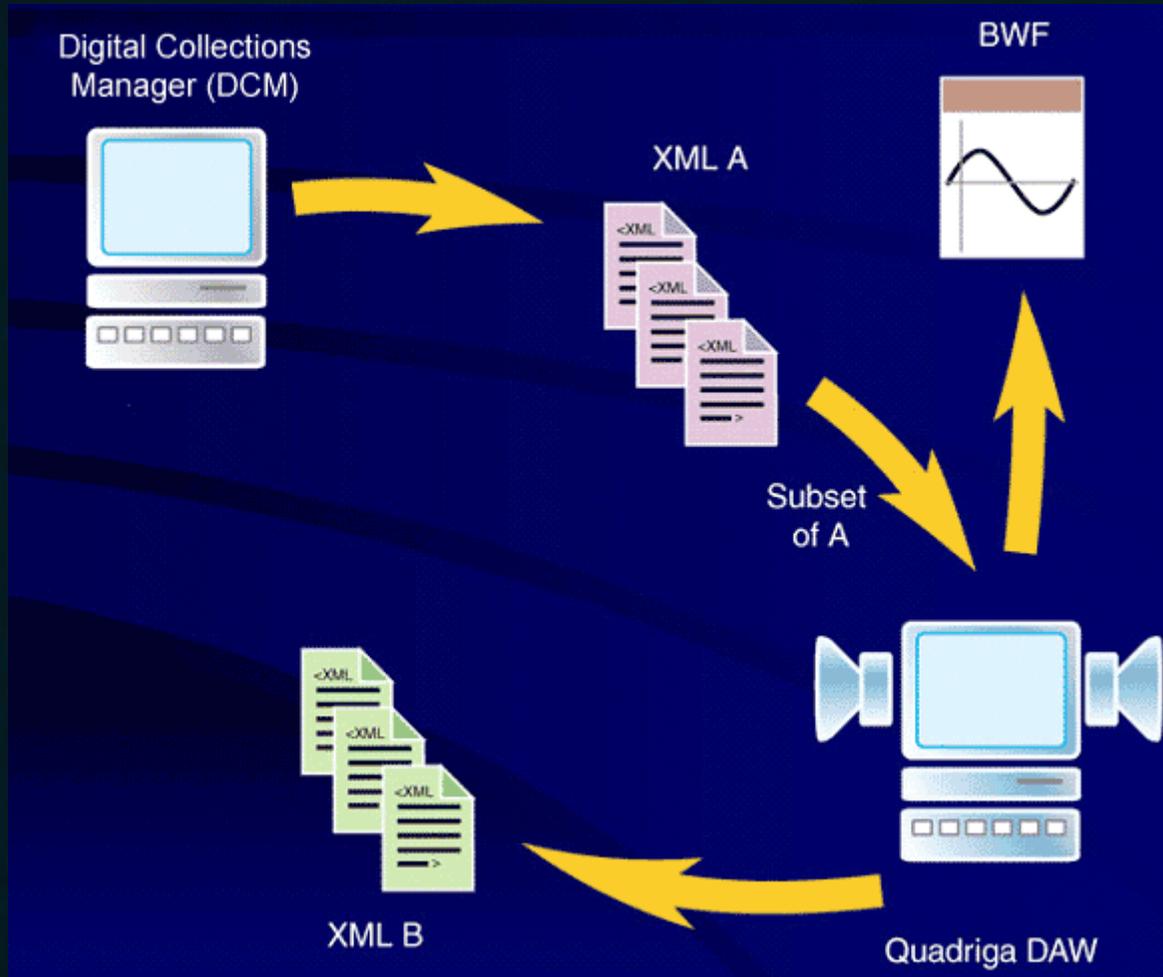


International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



### 3 *Create Digital Content*

Incorporation of metadata into BWF

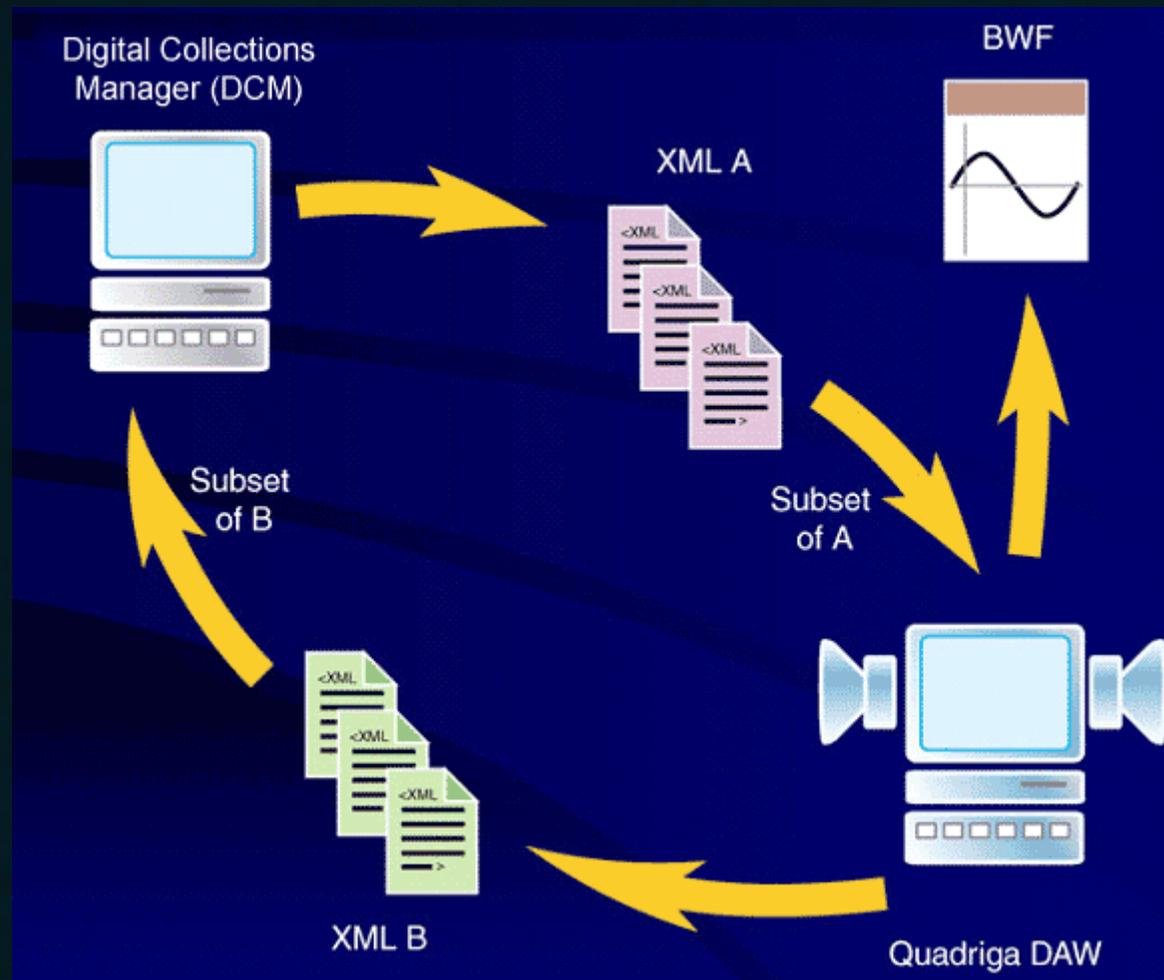


International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



### 3 *Create Digital Content*

Incorporation of metadata into BWF

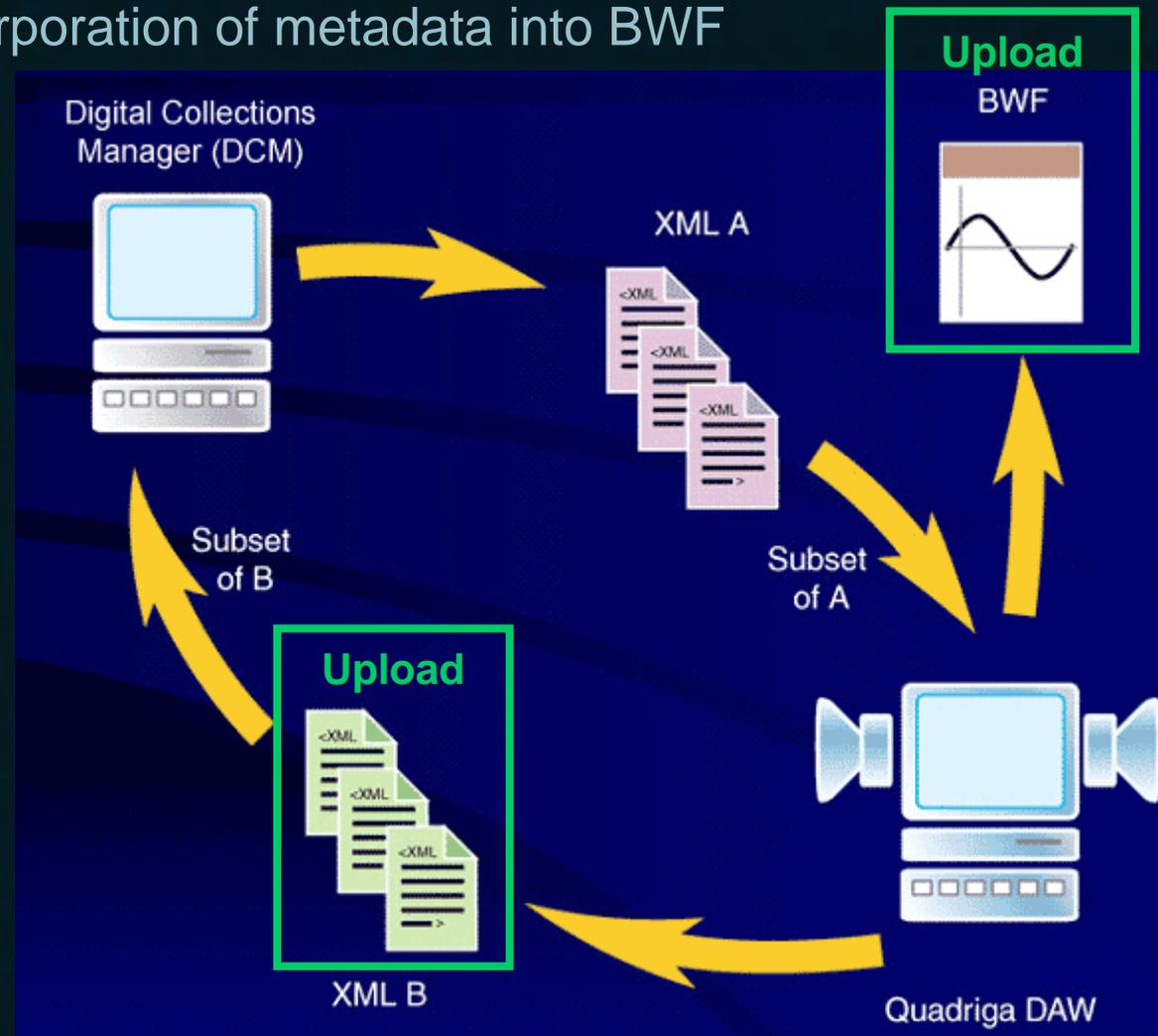


International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



### 3 *Create Digital Content*

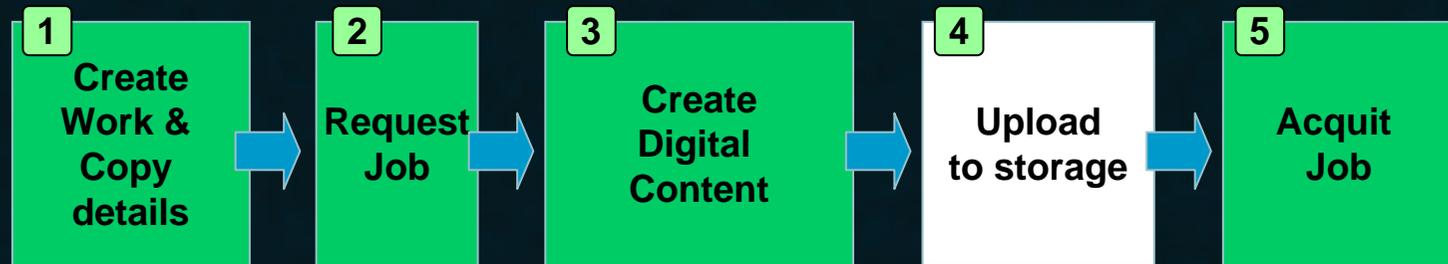
Incorporation of metadata into BWF



International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



# 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

International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



NATIONAL  
LIBRARY OF AUSTRALIA



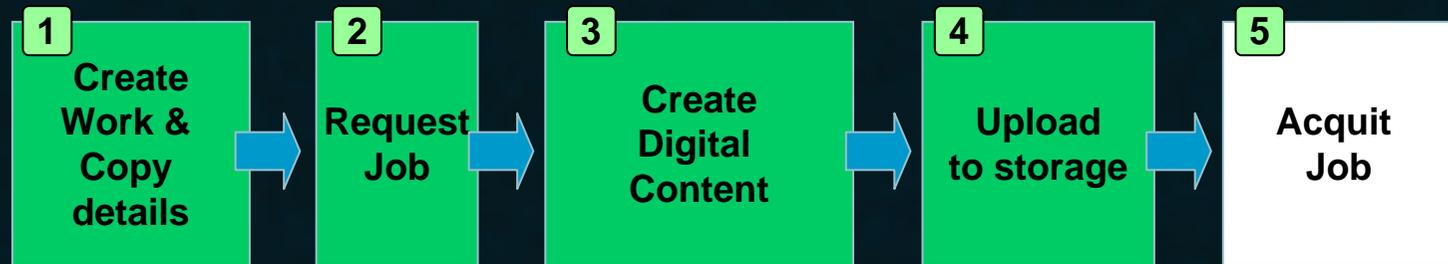
# Digitisation Workflow



- 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

International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany

# Digitisation Workflow



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

International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany

# Examples

## Image – Master copy detail

**Description** **History**

Find a work > [Result set](#) > [Work](#) > Copy    

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

<b>Copy ID:</b> nla.pic-an24283499-m-v1	
<b>Work ID:</b> <a href="#">nla.pic-an24283499</a>	
<b>File name:</b> nla.pic-an000024283499-m.tif	
<b>Access:</b> Unrestricted	
<b>File format:</b> image/tiff (TIFF file)	
<b>File size:</b> 106,144,566 bytes (101.2 MB)	
<b>Encoding:</b> MAC	
<b>Compression:</b> None	
<b>Checksum:</b> 1f2bf0c2d79dbd49ef686352988e64e8 (MD5)	
<b>Resolution:</b> 300x300 dpi	
<b>Bit depth:</b> 24 bits	
<b>Colour space:</b> RGB	
<b>Orientation:</b> Portrait	
<b>Dimensions:</b> 5,542 x 6,384 pixels	
<b>Date created:</b> 3 Feb 2003	
<b>Source copy:</b> <a href="#">nla.pic-an24283499-o-v1</a>	
<b>Agent:</b> Imaging Services (job no. <a href="#">5331</a> )	

[Co-master available](#)

Material type: Image; Copy Type: Digital; Copy Role: Master; Carrier: Online; Verson: 1;  
Created: 3 Feb 2003 (DCM) ; Last updated: 5 Feb 2003 ( )

International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany

# Examples

## Audio – Master copy detail

**Description** **Versions** **History**

[Find a work](#) > [Result set](#) > [Work](#) > [Copy](#) > [Copy](#) 

Displaying copy details for nla.oh-4599-0047-0001-m-v2:

**Copy ID:** nla.oh-4599-0047-0001-m-v2  
**Work ID:** [nla.oh-4599-0047-0001](#)  
**File name:** nla.oh-004599-0047-0001-m.wav  
**Access:** Unrestricted  
**Track:** 2  
**Sound field:** Stereo  
**Duration:** 00:53:39  
**Duration type:** Input  
**File format:** audio/x-wav (Microsoft RIFF)  
**File size:** 618,469,204 bytes (589.8 MB)  
**Compression:** None  
**Checksum:** db981083084309cdd42ca83b6a767116 (MD5)  
**Bit depth:** 16 bits  
**Sampling rate:** 48000  
**Date created:** 22 Aug 2005  
**Source copy:** [nla.oh-4599-0047-0001-o-v1](#)  
**Agent:** SPATS (job no. [27013](#))

Material type: Sound; Copy Type: Digital; Copy Role: Master; Carrier: Online; Verson: 2;  
Created: 22 Aug 2005 (DCM) ; Last updated: 29 Aug 2005 (gclifton)

International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



# Examples

## Audio – Master process history

Description	Versions	History		
<a href="#">Find a work</a> > <a href="#">Result set</a> > <a href="#">Work</a> > <a href="#">Copy</a> > <a href="#">Copy</a> 				
Displaying process history for nla.oh-4599-0047-0001-m-v2:				
Process	Input copy	Output copy	Status	Date assigned
<a href="#">Replicate</a>	<a href="#">nla.oh-4599-0047-0001-o-v1</a>	nla.oh-4599-0047-0001-m-v2	Acquitted	22 Aug 2005
<a href="#">Ingest</a>	nla.oh-4599-0047-0001-m-v2		Acquitted	22 Aug 2005
<a href="#">Derive</a>	nla.oh-4599-0047-0001-m-v2	<a href="#">nla.oh-4599-0047-0001-rb4-v1</a>	Acquitted	22 Aug 2005
<a href="#">Derive</a>	nla.oh-4599-0047-0001-m-v2	<a href="#">nla.oh-4599-0047-0001-ra1-v1</a>	Acquitted	22 Aug 2005
<a href="#">Derive</a>	nla.oh-4599-0047-0001-m-v2	<a href="#">nla.oh-4599-0047-0001-rm-v1</a>	Acquitted	22 Aug 2005
<a href="#">Derive</a>	nla.oh-4599-0047-0001-m-v2	<a href="#">nla.oh-4599-0047-0001-rb2-v1</a>	Acquitted	22 Aug 2005
<a href="#">Derive</a>	nla.oh-4599-0047-0001-m-v2	<a href="#">nla.oh-4599-0047-0001-rb3-v1</a>	Acquitted	22 Aug 2005
<a href="#">Derive</a>	nla.oh-4599-0047-0001-m-v2	<a href="#">nla.oh-4599-0047-0001-sa1-v1</a>	Acquitted	22 Aug 2005
<a href="#">Derive</a>	nla.oh-4599-0047-0001-m-v2	<a href="#">nla.oh-4599-0047-0001-sb2-v1</a>	Acquitted	22 Aug 2005
<a href="#">Derive</a>	nla.oh-4599-0047-0001-m-v2	<a href="#">nla.oh-4599-0047-0001-sb1-v1</a>	Acquitted	22 Aug 2005
<a href="#">Derive</a>	nla.oh-4599-0047-0001-m-v2	<a href="#">nla.oh-4599-0047-0001-rb1-v1</a>	Acquitted	22 Aug 2005

International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany

# Examples

## Audio – Master process detail

**Description**

[Find a work](#) > [Result set](#) > [Work](#) > [Copy](#) > [Copy](#) > [Process](#)   

Displaying process details for the nominated process:

**Job number:** [27013](#)  
**Process type:** Replicate  
**Process role:** Primary  
**Input copy:** [nla.oh-4599-0047-0001-o-v1](#)  
**Output copy:** [nla.oh-4599-0047-0001-m-v2](#)  
**Status:** Acquitted  
**Date assigned:** 22 Aug 2005  
**Signal source:** [DAT-Inspector](#)

International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



# Web archiving: PANDAS

PANDORA  
Archive

PANDORA  
AUSTRALIA'S WEB ARCHIVE  
National Library of Australia and Partners

Search PANDORA Search help

Archiving web resources : issues for cultural heritage institutions : international conference

This title is no longer being archived. The [publisher's site](#) may provide more information.

[Site](#) archived 22 Feb 2005  
[Site](#) archived 12 Nov 2004

To cite this page  
Create Persistent URL

[Copyright Warning](#) | [About](#)

Log Off Help Advanced Title View Panda

Basic 45920 Archiving web resources : issues for cultural heritage Institutions : international conference

**Kinetics No.** 26513497  
**Local Reference**  
**Publisher's URL** <http://www.nla.gov.au/webarchiving/index.html>

**Format** Mono  
**Subjects** Computers & Internet  
**Registration** 12 November 2004  
**Status** Ceased  
**Owner** NLA (Paul Koerbin)

**Indexing Agency**  
**Collections**  
**Gather Settings** Gathered Never

Close

Standard User (AgAdmin) (PADISK)  
Version 2.2.1

Local intranet

International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



PANDAS  
(PANDORA Digital  
Archiving System)



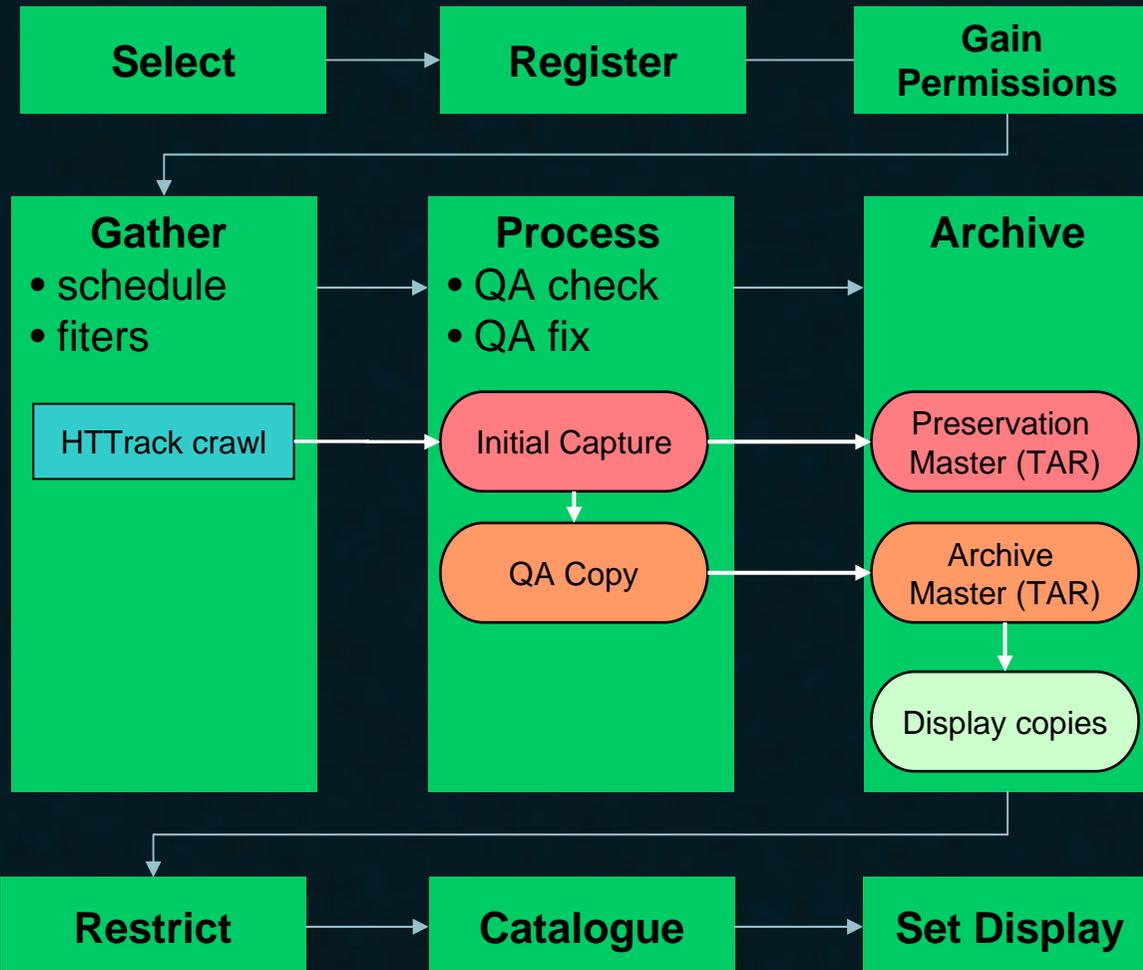
# Web Archiving: PANDAS

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



International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany

# Web Archiving Workflow



International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



# *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

International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany

# *Future enhancements*

- Digital Collections Manager
  - Additional formats (e.g. PDF)
  - Rights management
- PANDAS
  - Batch loading
  - IA Heritrix crawler and ARC format

International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany

# *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

International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany



# Conclusion

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

International Conference  
on Preservation of  
Digital Objects  
15 - 16 September 2005,  
Goettingen, Germany

