

Digital collecting and Prometheus

Update 2012



Digital Preservation responsibilities

Preservation of the Library's digital collections involves three main goals:

- maintaining access to reliable data at the **bit-stream level**
- maintaining access to **content encoded within the bit-streams**
- maintaining access to **the intended and available meaning of that content**

While specific preservation activities may focus on one or more of these goals, the **Library's preservation responsibility is only fulfilled when all three goals have been adequately addressed.**

This responsibility applies across all digital collections, **subject to the curatorial and policy decisions** for specific groups of digital objects.

Prometheus: a pre-history

The Digital Preservation Workflow Project (DPW) was developed as a mechanism to:

- **support** the concept of providing digital collection support (at the point of collection)
- **migrate** acquired digital content off physical format carriers before the carriers either failed, or possibly the e-mail attached PDF is stored elsewhere
- the hardware and software required to access the physical format carriers' content was no longer supported, e.g. floppy disks

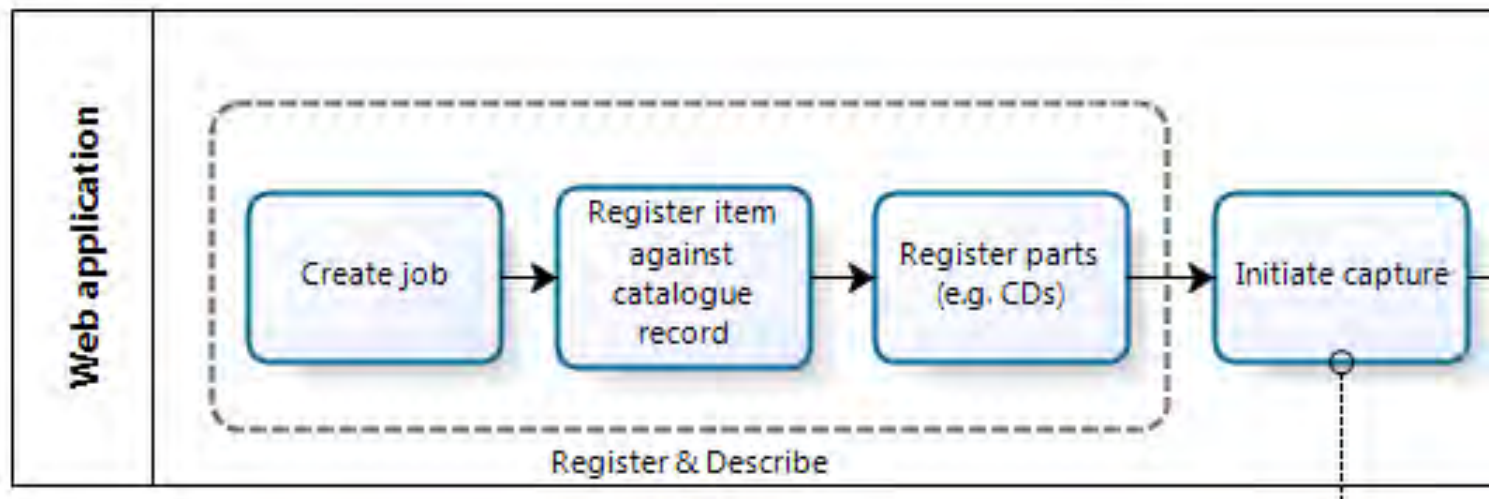


Workflow

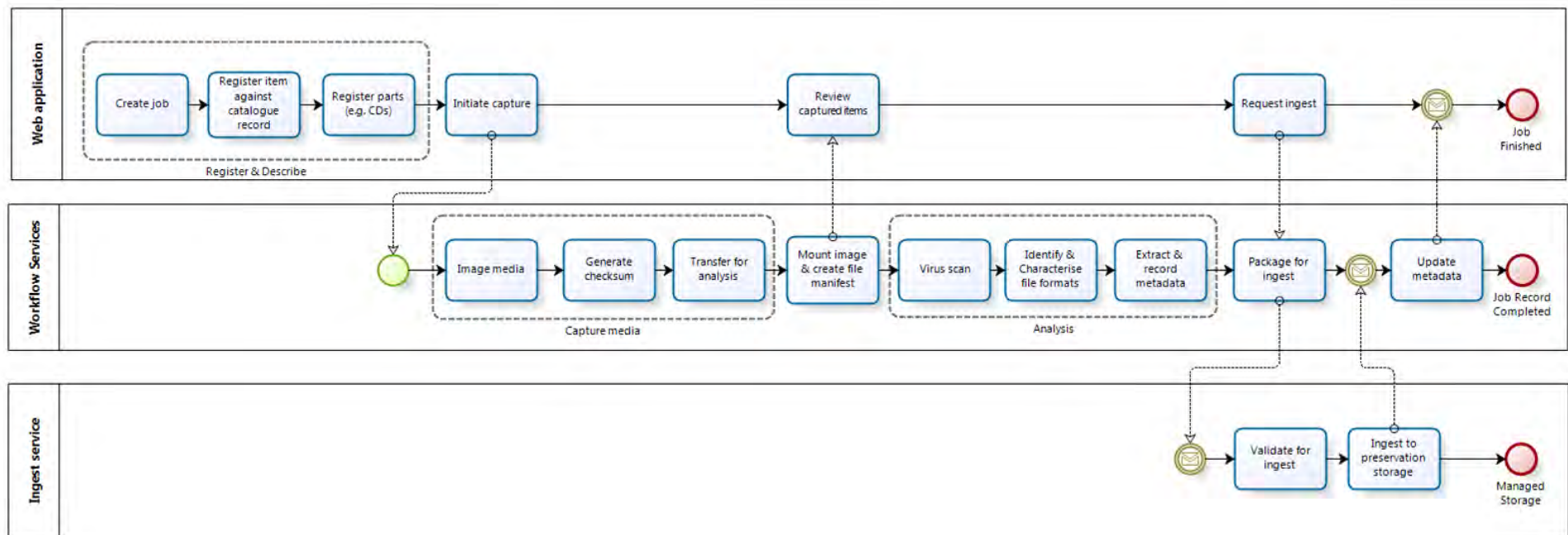
Digital Preservation utilised a number of:

- newly developing **software tools**
- **workflow processes**
- **digital preservation software tools**

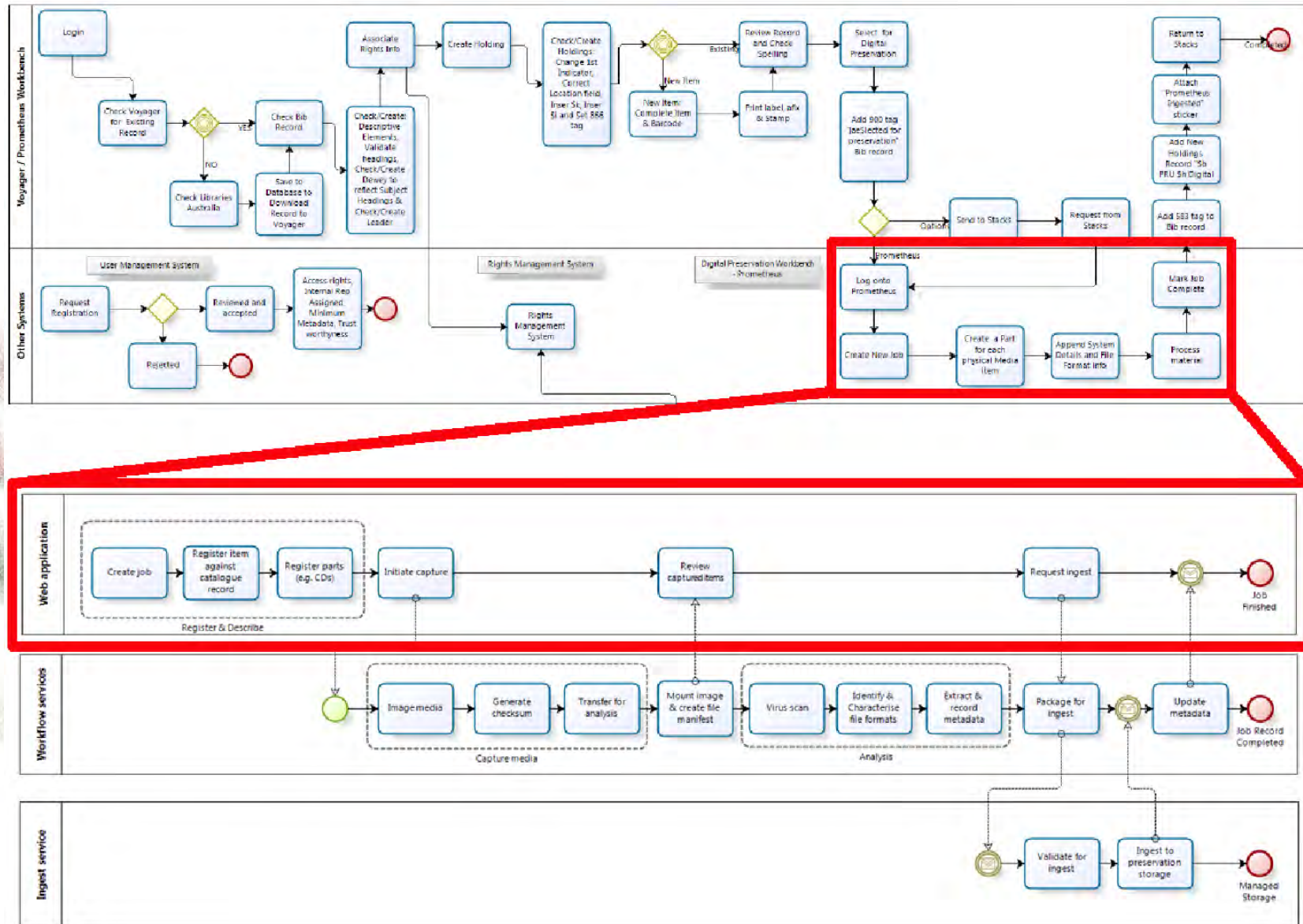
These were used to prototype a **generic framework** that could be extended to incorporate other future digital workflows.



Workflow system



Workflow system in context



Single electronic publication example

electronic publication such as ePub, PDF, doc, xls, MS pub, jpg/jpeg, etc.

The screenshot displays the PROMETHEUS digital preservation workbench interface. The header includes the PROMETHEUS logo, the National Library of Australia logo, and navigation links for Login, About, and Help. A search bar is also present. The main content area is titled 'Soldier of the Brell' and features a 'Media' tab and a 'Browse Captured Files' button. A table lists the media items, showing 'Soldier of the Brell' as a completed file. Below the table, there are sections for Annotations, Part PI, File Format, and System Details. On the right side, there are tabs for Job Details, Bibliographic Information, Copyright Information, and Notes & Metadata. The Bibliographic Information section is expanded, showing details such as Bib ID, Holdings, Author, Title, Description, System Details, and Notes.

Part Name <small>See naming</small>	Status	Media Type
▼ Soldier of the Brell	Completed	File

Annotations

Part PI	nla.dp-n861536
File Format	EPUB
System Details	epub compatible device

Bibliographic Information

Bib ID	4941907
Holdings	Default Digital Holdings nla.dp-n82
Author	Scholes, David K.
Title	Soldier of the Brell [electronic resource] / David K. Scholes.
Description	100 p.
System Details	System requirements: epub compatible device.
Notes	

Software analysis tools

Analysis tools implemented in the Prometheus workflow were in their early development. These have since been upgraded and refined.

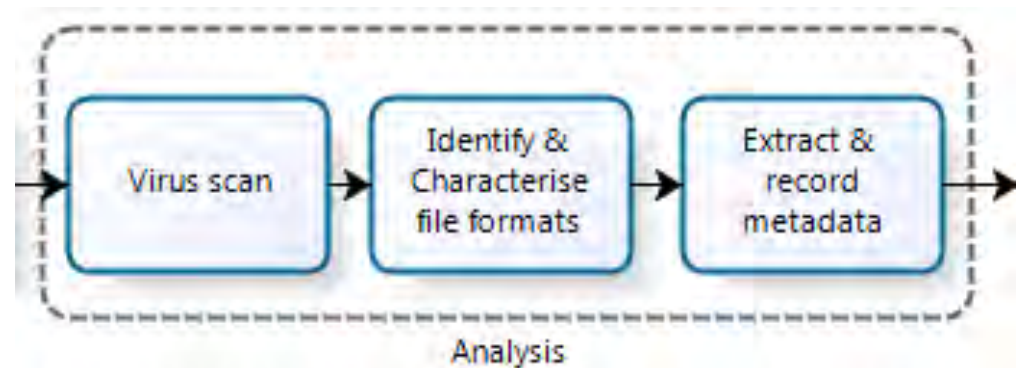
They are essential for:

- aiding in the **identification and verification** of digital objects
- **highlighting inherent potential risks** of long term access to digital objects

There is some crossover between the tools, however different output schemas are used.

Tools in use

- **virus scanning** - ClamAV
- **file format identification** – DROID (using PRONOM file format signatures)
- **file characterisation** – JHOVE file format specification validation (limited number of formats)
- **metadata extraction** – NLNZ Metadata Extractor



2012 update

- implementation and **refinement of workflow** including the ability to Restart a Task that was in stopped in an ERROR state
- implement ability to **process and ingest VCDs**
- improve **stability of various web services**
- **optimisation of cataloguing service** for recording technical and preservation metadata
- **optimisation of the search functionality** and speed
- implementation of **DCM Downloader tool** to enable access to content within a former 'Dark Archive'
- **optimise response of user interface** and Ingest service
- improve **mounting and reading of DVD video file systems**
- identify and quickly **rectify disc issues in Analysis**

DCM Downloader tool

DCM Bulk File Stager and Downloader

Step 1: Enter a list of PI's each on a new line

nla.dp-n2521223|

Note: This is a list of the PI's only. For example: nla.int-ex110-s15

Step 2: Select Agent Working Directory

digipres ▼

Note: The images are copied into a temp folder under the chosen working directy.

Step 3: Select Copy type

master ▼

Note: The most recent version is downloaded.

Submit Query

Lab environment compilation

- to provide access to applications with specific **hardware and software dependencies** for physical format carriers
- **access to software and their associated dependencies** e.g. fonts (to prevent pagination changes for citations in manuscripts)
- building **View Paths to create future emulated environments** with useful applications (as per above), beyond their common use (typically used for games)



Other activities and outcomes

- develop the **Mediapedia** knowledgebase on a wide variety of physical media and their dependencies (2009 – ongoing)
- producing the **Collection Profiling report** (2009)
- producing the **Pandora File Format Analysis report** for the International Internet Preservation Consortium - IIPC (2010)
- producing a **Brief Survey of Contemporary Digital Repository Projects** for a NSLA presentation (2010)
- **evaluating** two existing digital repositories (Rosetta, RODA) functionality (2011)
- compiling lists of **potential software tools** for DLIR, including new tools for **Curatorial Assessment, Policy, Rights, Q&A, Analysis, Metadata, Processing, Forensics, Delivery and Access** (2007 – 2012)
- updating **functional and access requirements** for DLIR
- contributing to the **e-Publications Workflow Scoping project** (2011)

Digital collecting and Prometheus

Session 2



Overview

- **workflows for content types**
 - single electronic publications
 - adding a serial to an existing series
 - mixed collection items
- **support resources**
- **follow-up**
 - hands on sessions
 - reviewing and updating online documentation
 - reviewing and updating assessment software tools
 - developing a ongoing user reference group



Workflows for content types

Single electronic publication

e.g. electronic publication such as ePub, PDF, doc, MS pub, jpg/jpeg, etc.



The screenshot displays the PROMETHEUS digital preservation workbench interface. The header includes the PROMETHEUS logo, the National Library of Australia logo, and navigation links for Login, About, and Help. A search bar is also present.

The main content area is titled "Soldier of the Brell". It features a "Media" tab and a "Browse Captured Files" button. Below this, a table lists the job details:

Part Name	Status	Media Type
Soldier of the Brell	Completed	File

Below the table, there are sections for "Annotations", "Part PI", "File Format", and "System Details".

On the right side, there are tabs for "Job Details", "Bibliographic Information", "Copyright Information", and "Notes & Metadata". The "Job Details" tab is active, showing the following information:

Job Number	Owner	Date Added	Last Update	Priority	Status
10466	John Maguire	17 Aug 10, 05:05 PM	17 Aug 10, 05:05 PM	Medium	Finished

Workflows for content types

Single electronic publication

- these single files can be selected using the 'File' media option on the right
- This workflow is documented in Section 9.2-Capturing non-Optical Media of the on-line Workflow Help



Workflows for content types

Adding a serial to an existing series

When a Job for a series has already been created, the user is offered the choice of **adding to the existing Job** or **creating an entirely new Job**.



Create or Find Job

Jobs

Jobs already exist for this catalogue record and holdings.

You may *create a new job* for this catalogue record and holdings or *add to an existing job* below.

Job No.	Name	Last Updated	
742	Cemeteries of southern Tasmania. Vol. IX - Oatlands district CD 2	30 October 08, 12:20 PM	Add to this job

Create new job

Enter name of new job:

[Start again](#) [Cancel](#) [Create New Job](#)

Workflows for content types

Mixed collection items

ILMS Bib: vn4406497

Suddenly single : a new musical / book, music & lyrics by Peter Pinne & Paul Dellit.

ILMS Holdings: MUS Nmt 782.14 P656

1 score (112 leaves in various pagings) : ill., col. ports ; 30 cm. + 1 DVD

<http://www.suddenlysingletthemusical.com/Home.html>

Includes the script, score, DVD, program, flyer and some press for the musical, which took place at the Athenaeum Theatre, Melbourne, Monday December 10, 2007.

Workflows for content types

Unable to read legacy media

If **capturing a Part is unsuccessful**, due to issues reading, or unavailability of the legacy physical format carrier – the user can record that failure of the physical carrier within it's **Part**, as a **Stub**. This enables documenting that this Part existed within the series. (The user will be prompted to enter a reason for not capturing the data.)



Mark Part as Stub

Are you sure you want to mark the part named 'small UDF DVD' as a stub ?

Note: If the part has a status of 'Completed' then this process may take some time depending on the amount of metadata already recorded against this part.

Processing commences, but cannot be Completed due to errors in the media.

Reason

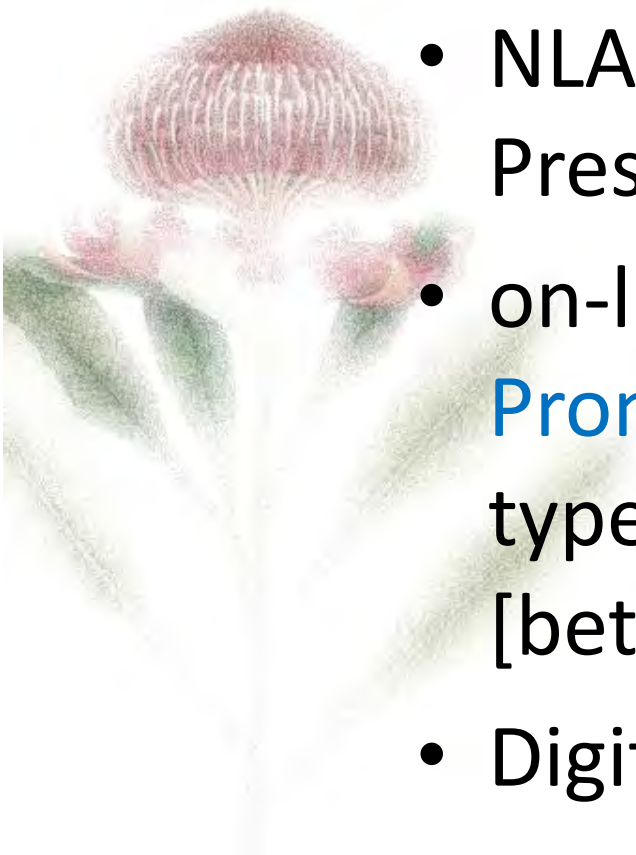
Mark as StubCancel

Providing digital collection support

- collection support in the form of direct enquiries

digipres@nla.gov.au

- NLA [Service Desk](#) – Digital Preservation Service Group
- on-line references such as the [Prometheus](#), [Mediapedias](#) for carrier types and the [Knowledge Bases](#) [beta]
- Digital Preservation - [OurWiki](#)



NLA Service Desk

- Digital Preservation Service Group
 - Digital Preservation Enquiry or
 - Prometheus application support or
 - Prometheus Documentation or
 - Prometheus training



National Library of Australia Service Desk

[Logout](#) [New Ticket](#) [MyTickets](#) [CompanyTickets](#) [Search](#) [FAQ-Area](#) [Preferences](#)

new ticket

Options

Type: Request ▾

Service Group: Digital Preservation ▾

Service: Digital Preservation ▾

Subject:

Text: **B** *I* U ~~ABC~~

Prometheus on-line resources

- Overview
- Workflow
- Cataloguing procedures (various)
- Media
- File formats
- System details
- Error messages
- Glossary
- FAQ
- About



NATIONAL
LIBRARY
OF AUSTRALIA

[Overview](#) | [Workflow](#) | [Cataloguing Procedures](#) | [Media](#) | [File Formats](#) | [System Details](#) | [Error Messages](#) | [Glossary](#) | [FAQ](#) | [About](#)

Overview – conceptual framework

ILMS Bibliographic Record

e.g. **4508277**

Great Western Highway upgrade:
Bullaburra West: Genevieve Road,
Bullaburra to Tableland Road,
Wentworth Falls

ILMS Holding Items – physical object
(s) including digital media carriers

e.g. **Nmt 9018**

Great Western Highway upgrade:
Bullaburra West: Genevieve Road,
Bullaburra to Tableland Road,
Wentworth Falls Disc 1 of 1

Prometheus Job

e.g. **1543, PI: nla.dp-n27166**

Great Western Highway
upgrade: Bullaburra West:
Genevieve Road, Bullaburra to
Tableland Road, Wentworth
Falls

Prometheus Part

e.g. **PI: nla.dp-n27168**

Great Western Highway upgrade:
Bullaburra West: Genevieve Road,
Bullaburra to Tableland Road,
Wentworth Falls Disc 1 of 1



[Overview](#) | [Workflow](#) | [Cataloguing Procedures](#) | [Media](#) | [File Formats](#) | [System Details](#) | [Error Messages](#) | [Glossary](#) | [FAQ](#) | [About](#)

Workflow

PROMETHEUS digital preservation workbench

NATIONAL LIBRARY OF AUSTRALIA

Logout About Help

My Jobs

My Jobs Parish maps, Hay Office set: Vol. PMAHY01, 15/11/2000

Media Browse Captured Files

Add media eg. CD, disk, software file...

Part Name	Status	Media Type
Town of Bairnald South to Town of Wombouta East, County of Boyd, Parish of Boyd to County of Cadel, Parish of Benarca. v2.1.2 Dec 1 of 8	Ready	cd-rom

Annotations:

System Details IBM PC or compatible, 200 MHz; 32 Mb RAM; Windows 95 or NT; Parish Map Viewer, MrSID viewer; 17 in. screen, 1024 x 768 resolution, 10x CD-ROM drive.

File Format File types include: .sid image files and MrSID system files .db, PmapSid.PX, PmapSid.XGO.

Bibliographic Information

Bib ID 3307591 [Change catalogue record](#)

Holdings (Copy 1) MAP film 3047 nla.dp-n41027 [Change holdings](#)

Author Land Information Centre (N.S.W.)

Title Parish maps, Hay Office set [cartographic material] : Town of

Step 6.9

The Status of the Part will then change from:

"Ready" to "Initialising"

Add media eg. CD, disk, network file...

Part Name	Status	Media Type
Solo piano 2	Initialising	CD/DVD-ROM

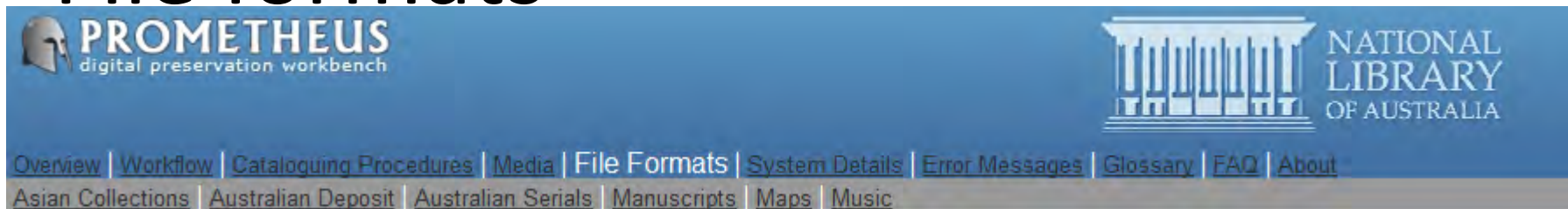
Capture Edit

Cataloguing procedures (various)

- Accessioning
- Appraisal
- Materials preparation
- Prometheus workflow
- Naming conventions
 - Job
 - Part
 - Stub
- Documentation finalisation
- Summary steps
- Glossary
- Do's and don'ts
- Policy links




File formats



Previously suggested file format extensions of interest to various collection areas

✓ - this file format is known by the NLA Knowledge Base [accessibility yet to be confirmed]

⚠ - this file format is currently unknown by our format identification tools and should be noted within the Prometheus File Format Annotations to enable further investigation in regards to long term support and access.



Home **Software** **File formats** **Hardware** **Physical carrier**

File formats

The Digital Preservation Knowledgebase currently holds information about file formats and will be increasingly expanded to hold information on software hardware and dependencies and limitations.

Included in the Knowledgebase are entries (from a shortlist of file formats), as selected by the Library's Collection Areas, based on commonly found formats that have already been ingested into Prometheus, appear in the collection or are likely to be collected by the Library in the near future. The Knowledgebase contains information about these file formats including a general description,

FileInfo.com

File Extension Search

Advanced Search

Browse Alphabetically ▶

Ads by Google

[MP3 to Wav](#)
[Wav MP3](#)
[WMA to Wav](#)


Common File Types

[Text Files](#)
[Data Files](#)
[Audio Files](#)
[Video Files](#)
[3D Image Files](#)
[Raster Image Files](#)
[Vector Image Files](#)
[Page Layout Files](#)
[Spreadsheet Files](#)

AVS Video Converter

www.avs4you.com

DVD AVI VOB MPEG MP4 MOV 3GP XviD
WMV MKV FLV SWF MOD TOD MTS M2TS
AVCHD H.264 DVR-MS M4V PPT TS...


Free Download 

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

.WAV File Extension

2 File Associations

TIP: [Run a free scan for .WAV file extension errors.](#)



File Type

Developer

Popularity

Category

WAVE Audio File

Microsoft

★★★★★

Audio Files

Standard digital audio file format used for storing waveform data; allows audio recordings to be saved with different sampling rates and bitrates; often saved in a 44.1 KHz, 16-bit, stereo format, which is the standard format used for CD audio.

Exif Audio

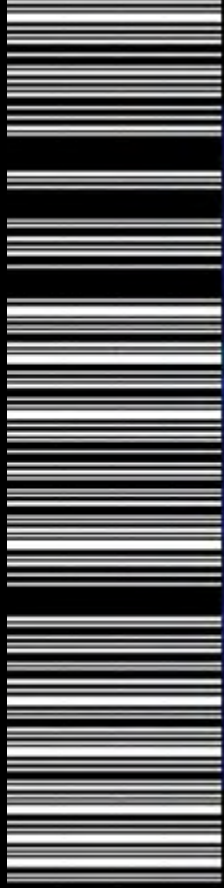
Exif Compressed Image

Follow up

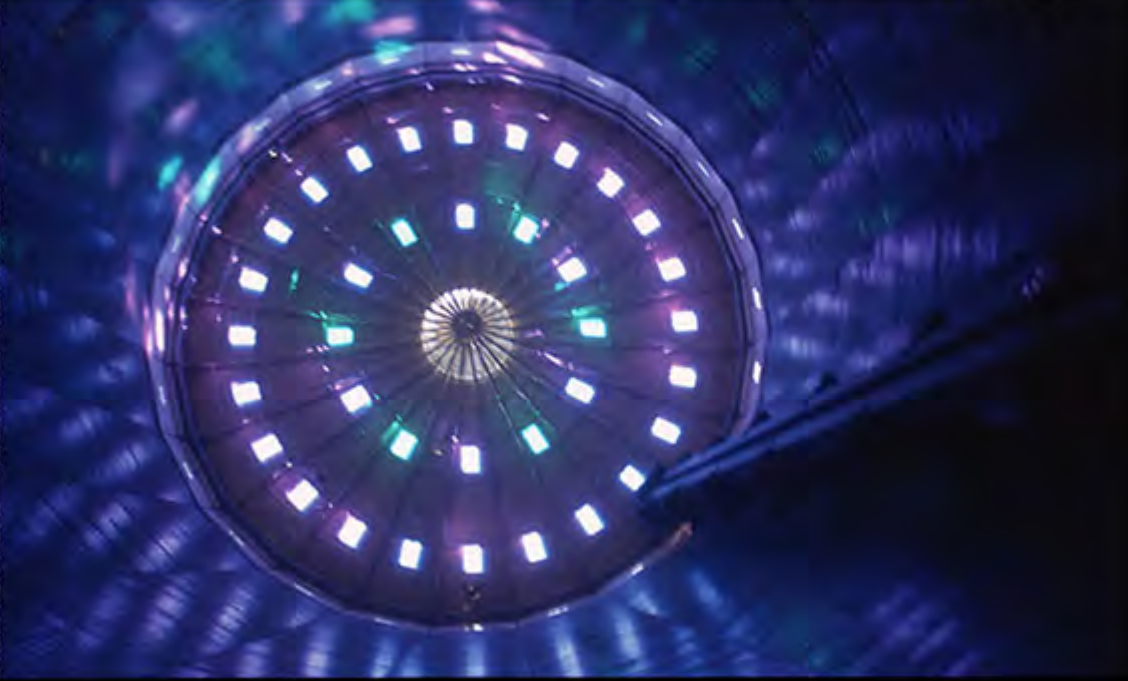


- **workflows for content types**
 - single electronic publications
 - adding a serial to an existing series
 - mixed collection items
- **support resources**
- **follow-up**
 - hands on sessions
 - reviewing and updating online documentation
 - reviewing and updating assessment software tools
 - developing a on-going user reference group





Managing Collections in a Digital Age



Douglas Elford