



NATIONAL ARCHIVES OF AUSTRALIA

# Digital Preservation at the NAA

## DPC Audiovisual Working Group

Nathan Andrews



Australian Government  
National Archives of Australia

# Acknowledgement of Country

The National Archives of Australia acknowledges the traditional owners and custodians of Country throughout Australia and acknowledges their continuing connection to land, sea and community. We pay our respects to the people, their cultures and Elders past, present and emerging.

# Overview

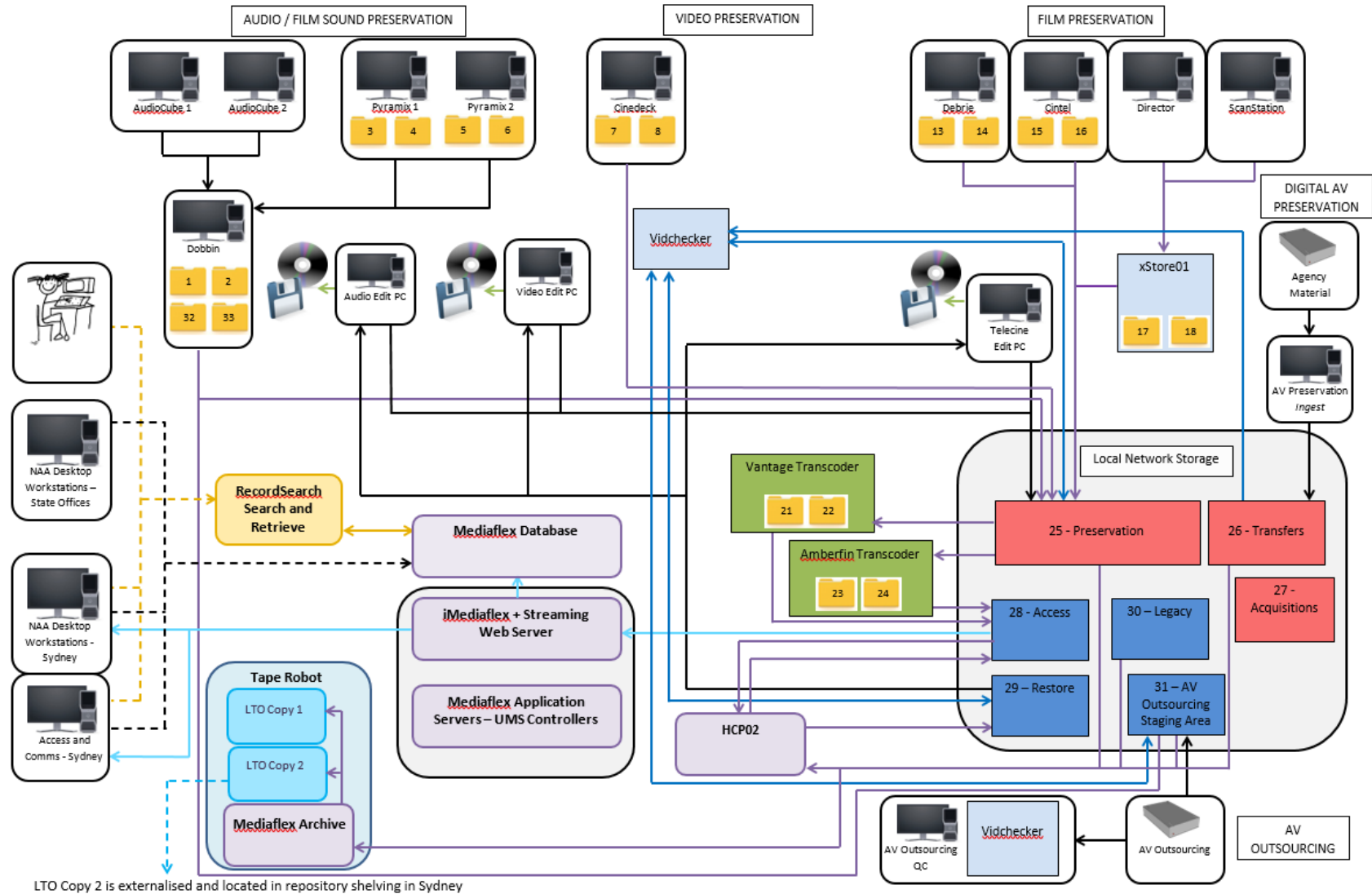
Today we will look at:

- A brief history of Digital Preservation at the National Archives of Australia
- Our digital ecosystem for AV Preservation
- Our system - Mediaflex
- How we do digital preservation for our AV collection
- Our challenges

# History of Digital AV Preservation at NAA

- 2001 – First digitisation of audio material. Standard based on forthcoming IASA TC04 - *Guidelines in the Production and Preservation of Digital Audio Objects*.
- 2002 – NAA develop "An Approach to the Preservation of Digital Records"
- 2002-2008? – NAA Digital Preservation Project creates the Digital Archive in Canberra.
- 2005-2008 – Pilot projects to put AV in Digital Archive
- 2012 – Digital AV to be managed in a new Audiovisual Archival Management System and an AV Digital Archive in Sydney
- 2005-2014 – Digital preservation principles and fixity managed manually using checksum tools.
- 2014 – Mediaflex implemented and workflows configured based on established digital preservation principles – fixity checks and metadata harvesting.
- 2014 – Mediaflex Archive LTO tape library chosen as AV Digital Archive. Redundancy managed through 3 copies on different brands of LTO stock
- 2019 – Hitachi Content Platform (HCP) S3 object store implemented as 2nd storage technology for AV digital preservation storage alongside MFX LTO tape library.

# Our Digital Ecosystem



# Mediaflex

The National Archives' AV collection is managed by Mediaflex – developed by TransMedia Dynamics (TMD)

Mediaflex:







- Gives us descriptive and archival control of our AV collection
- Provides control and asset management of the digital collection.
- Maintains the relationships between assets, eg, when digitisation produces digital copies of physical source items.
- Has the ability to design and configure workflows Mediaflex – these are used for the vast majority of what we do with the AV collection.
- Provides consistency and creates a reportable audit history of work and actions performed on collection items.



# File Fixity

Fixity is confirmed using MD5 checksums created and validated by Mediaflex workflows

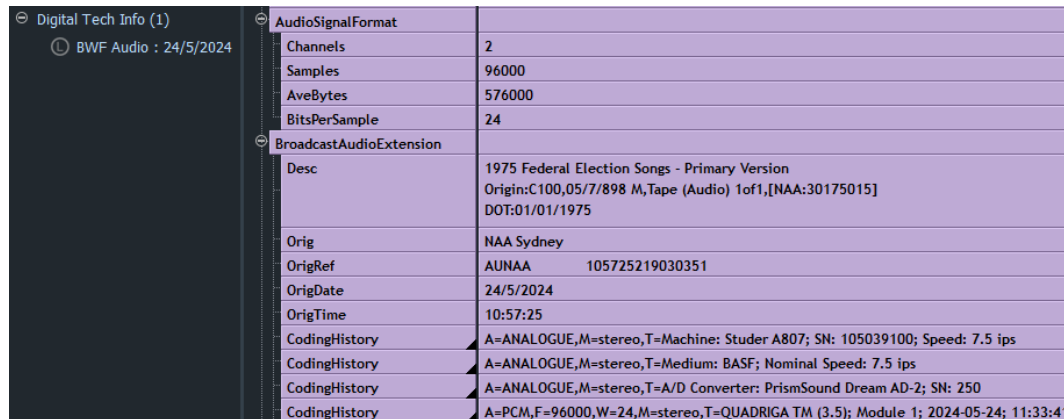
- We receive checksums from vendors and agencies for outsourced digitisation and born-digital agency transfers – these are imported into Mediaflex and validated by Mediaflex workflows
- File moves are performed by Mediaflex workflows – all moves are validated by checksum validation
- Any files moves prior to ingesting the file into Mediaflex (for example, from hard drives) are validated manually using tools such as HashCheck or TeraCopy

Full Title	Eps#	Icon	Format	Filename	Checksum	Checksum Type
World of Music / Series 82 / Sydney Symphony Orchestra: M...	8		JPEG 2000 MXF	C475_ARCH 09-2927_454269.mxf	4CC904121AC9066CDF86976CB92FF068	MD5
In The Wild With Harry Butler / Series 3 / North West Quee...	6		Digital Picture eXchange	C475_ARCH 92F-0152_509021.tar	068005DE169DCD6EBA5C77A65FCFDF3F	MD5
World of Music / Series 82 / Dvorak Violin Concerto In A Min...	4		JPEG 2000 MXF	C475_ARCH 09-2930_454327.mxf	161D53778BCA62B6FBF5BF7E5F49FC7E	MD5
Home Grown: Gardening with Graham and Sandra Ross / Se...	36		JPEG 2000 MXF	C475_ARCH 09-2934_454249.mxf	A541E6CD35DB1E7E206767EAF3CC8298	MD5
In The Wild With Harry Butler / Series 1 / Barrow Island - Pa...	5		PCM Waveform Audio	C475_ARCH 92F-0147A_509344.wav	A9D8D3F96376C398A6278CEB89C6889A	MD5
1975 Federal Election Songs			PCM Waveform Audio	C100_05-7-898 M_509428.wav	DF9112EB02D61BCB8EA84FD3EA1E63D1	MD5
World of Music / Series 81 / Sydney Symphony Orchestra Co...	11		JPEG 2000 MXF	C475_ARCH 09-2929_454217.mxf	5B27D9616A7012E6491D3A2AA949D96D	MD5

# Metadata

Workflows in Mediaflex harvest technical metadata for all digital AV in the collection

- Mediaflex uses MediaInfo as a third-party tool and combines this with metadata harvested from files directly or transformed from side-car xml files.
- Metadata is saved as xml and we have schemas based on format and digitisation system.
- Audio has a Dobbin-specific schema which contains basic metadata combined with quality events and parameters captured by Dobbin. Mediaflex metadata is also embedded in BWF by Dobbin.
- Legacy video has a Samma-specific schema which combines MediaInfo metadata with metadata from the Samma xml side-car file.
- All other video has a general video schema containing a subset of key MediaInfo fields
- Agency transfers has a MediaInfo schema that captures all metadata that MediaInfo is able to capture

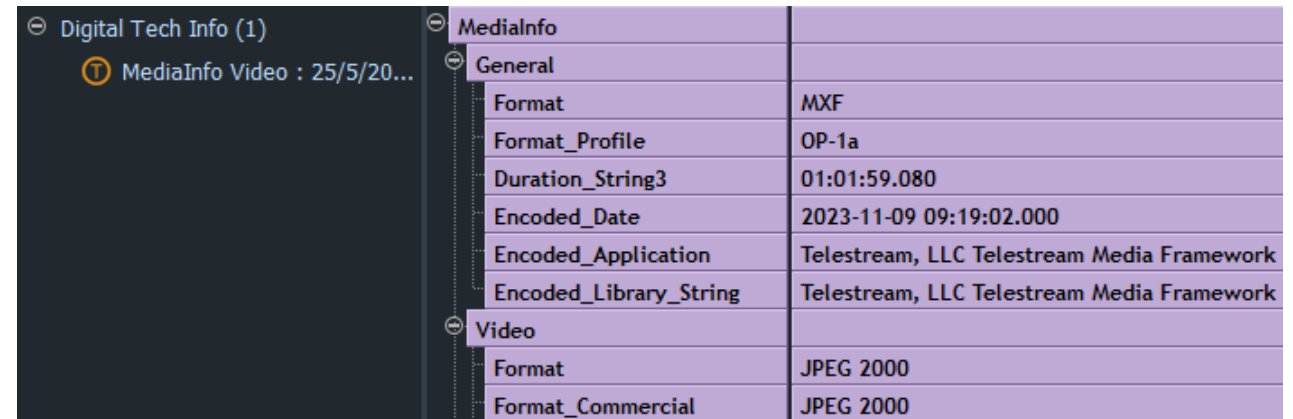


Digital Tech Info (1)  
BWF Audio : 24/5/2024

AudioSignalFormat	
Channels	2
Samples	96000
AveBytes	576000
BitsPerSample	24

BroadcastAudioExtension	
Desc	1975 Federal Election Songs - Primary Version Origin:C100,05/7/898 M,Tape (Audio) 1of1,[NAA:30175015] DOT:01/01/1975
Orig	NAA Sydney
OrigRef	AUNAA 105725219030351
OrigDate	24/5/2024
OrigTime	10:57:25
CodingHistory	A=ANALOGUE,M=stereo,T=Machine: Studer A807; SN: 105039100; Speed: 7.5 ips
CodingHistory	A=ANALOGUE,M=stereo,T=Medium: BASF; Nominal Speed: 7.5 ips
CodingHistory	A=ANALOGUE,M=stereo,T=A/D Converter: PrismSound Dream AD-2; SN: 250
CodingHistory	A=PCM,F=96000,W=24,M=stereo,T=QUADRIGA TM (3.5); Module 1; 2024-05-24; 11:33:41




Digital Tech Info (1)  
MediaInfo Video : 25/5/20...

MedialInfo	
General	
Format	MXF
Format_Profile	OP-1a
Duration_String3	01:01:59.080
Encoded_Date	2023-11-09 09:19:02.000
Encoded_Application	Telestream, LLC Telestream Media Framework
Encoded_Library_String	Telestream, LLC Telestream Media Framework
Video	
Format	JPEG 2000
Format_Commercial	JPEG 2000

# QC

Quality assurance and quality checks are performed to ensure that we are putting good files into our systems.

- Dobbin performs QC on all audio files, checking that technical parameters meet our standards
- Vidchecker performs QC on video files – currently only JPEG2000 files created through digitisation
- DPX sequences created from motion picture film are QC'd manually and we have recently procured GrayMeta to QC DPX sequences
- All files received from outsource vendors have a basic level of QA to ensure files are good before we accept them. This includes:
  - A virus scan
  - A technical metadata check using MDQC
  - Manual checksum validation
- Currently no QC occurs for agency transfers beyond checksum validation

Full Title	Eps#	Icon	Format	Filename	Assisted QC
Australia You're Standing In It / Series 1 / [28/09/1983]	1		JPEG 2000 MXF	B5082_CV83187_454355.mxf	Pass
Australia You're Standing In It / Series 1 / [05/10/1983]	2		JPEG 2000 MXF	B5082_CV83188_454429.mxf	Pass
Australia You're Standing In It / Series 1 / [19/10/1983]	4		JPEG 2000 MXF	B5082_CV83190_454411.mxf	Pass
Australia You're Standing In It / Series 1 / [12/10/1983]	3		JPEG 2000 MXF	B5082_CV83189_454435.mxf	Pass
Australia You're Standing In It / Series 1 / [02/11/1983]	6		JPEG 2000 MXF	B5082_CV83192_454409.mxf	Pass
A New World (For Sure) / The Seeding Time			JPEG 2000 MXF	B5082_CV84127_454459.mxf	Pass

# Storage

Preservation storage is a combination Hitachi HCP S3 object storage + LTO tape

- All files are written to the HCP and LTO tape in parallel
- The HCP replicates between two of the NAA's largest sites in Sydney and Canberra respectively - the replicated copy is not transparent to Mediaflex
- The HCP has its own file fixity and deduplication technology that is not transparent to Mediaflex
- File integrity and authenticity relies on a combination of storage-level (HCP) and application-level (Mediaflex) processes to ensure what went in is what comes out
- Additional redundancy is achieved by using different brands of LTO tape

# File Formats - Standards

- We have published our Born-Digital File Format Standards and Preservation Digitisation Standards and ensure alignment between these two standards
- Files created through digitisation are chosen in line with digital preservation good practice and current technology in the AV Preservation industry
- Preservation Digitisation Standards apply to in-house and outsourced digitisation, and agency digitisation of physical AV records
- Current file formats used are:
  - BWF for audio
  - JPEG2000 in an MXF wrapper for video
  - DPX for motion picture film, with a ProRes mezzanine copy
- Born-digital files created by agencies should conform with our Born-Digital File Format Standards, however agencies are advised not to convert files to our standard that are not originally in these formats
- The National Archives currently does not perform file format migrations for any of the digital collection, however Mediaflex integrates with third-party transcoders - we use Amberfin and Vantage

# File Formats – Risk Management

We have recently completed a file format risk assessment on all formats in the collection (AV and non-AV)

- Mediaflex is not a digital preservation system and so doesn't characterise file formats like DROID with PRONOM
- Mediaflex matches file formats using the extension but can harvest rich technical metadata to assess file formats
- We haven't started looking deeper into the file formats of our born-digital AV collection however initial investigation indicates the complexity of video files, having video and audio streams in many combinations within wrapper formats.
- Initial thoughts is that DROID and PROMON may be too simplified to properly assess digital video formats.
- Files created through digitisation are consistent and based on standards and so do not carry the same risk as born-digital files received from agencies.

# Challenges

- File Sizes
- DPX sequences
- File format characterisation
- LTO library connection interruptions
- Cyber Security impacts



**Australian Government**

---

**National Archives of Australia**

# Mediaflex – Title Search

Titles **General 1** > ↔ ⊕

Series	Title	Full Title	Medium	Form	Genre	Cultural Sta...	Master Title ...
	May Day Mel...	May Day Melbourne 1953	Moving Image				1297540

Versions **Default** > ↔ ⊕

Tree	Full Title	Version	Colour	Audio	Mo...	Year	Dur	Cultural Sta...	Version ...
SM	May Day Melbourne 1953	Primary Version	Black & White	Silent	1	May 1953	2:58		1297542

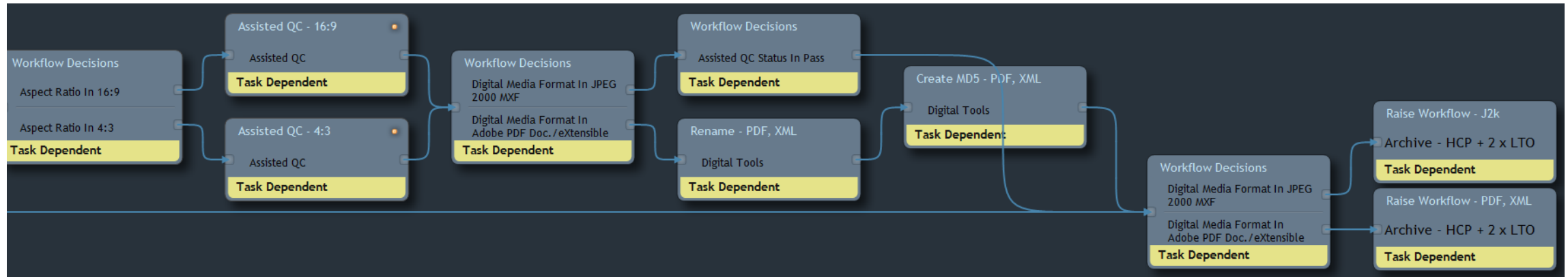
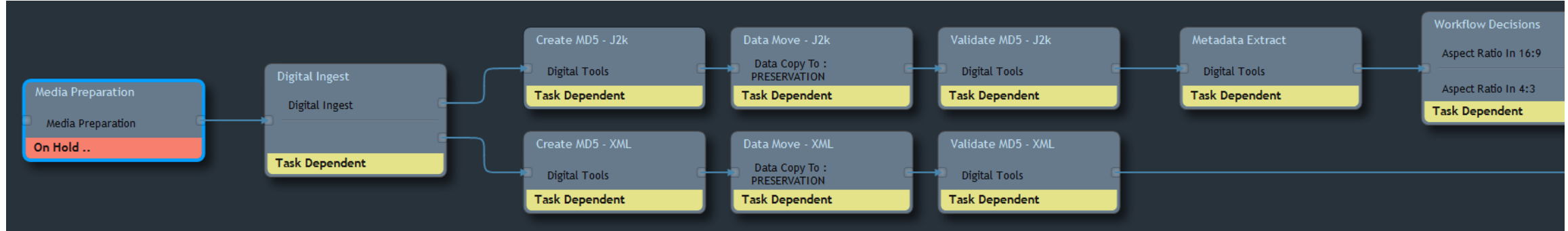
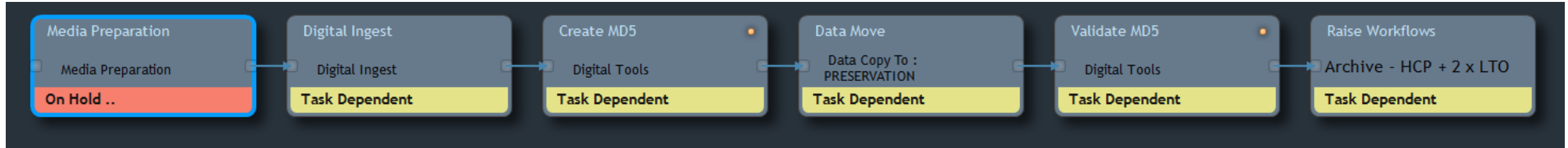
Media **Default** > ↔ ⊕

Full Title	E...	Icon	Format	Thumb	Filename	File Size	Checksum	Ctnr. ID
May Day Melbourne 1953			Apple QuickTime Movie		A8703_32-1-300 Volume 3_508966.mov	35.3 GB	25BBE05D6C81B896FAD2DD8FB9AE256D	HITACHI HCP - PRESER...
May Day Melbourne 1953			Apple QuickTime Movie [ARC]		A8703_32-1-300 Volume 3_508966.mov	35.3 GB	25BBE05D6C81B896FAD2DD8FB9AE256D	TMD_ARCHIVE
May Day Melbourne 1953			Apple QuickTime Movie [ARC]		A8703_32-1-300 Volume 3_508966.mov	35.3 GB	25BBE05D6C81B896FAD2DD8FB9AE256D	TMD_ARCHIVE
May Day Melbourne 1953			MPEG4		A8703_32-1-300 Volume 3_508968.mp4	155.9 MB	FB0D02685692696D71BF8645BB380109	ACCESS
May Day Melbourne 1953			MPEG4		A8703_32-1-300 Volume 3_508968.mp4	155.9 MB	FB0D02685692696D71BF8645BB380109	HITACHI HCP - ACCESS
May Day Melbourne 1953			16_N_B&W_D_Aw_A					10308587
May Day Melbourne 1953			ST_DBC_MP					10308592
May Day Melbourne 1953			JPEG 2000 MXF [ARC]		A8703_32-1-300_Volume_3.mxf	1.7 GB	ff4a7d01eea8ea99a480ad5bb5bab928	TMD_ARCHIVE

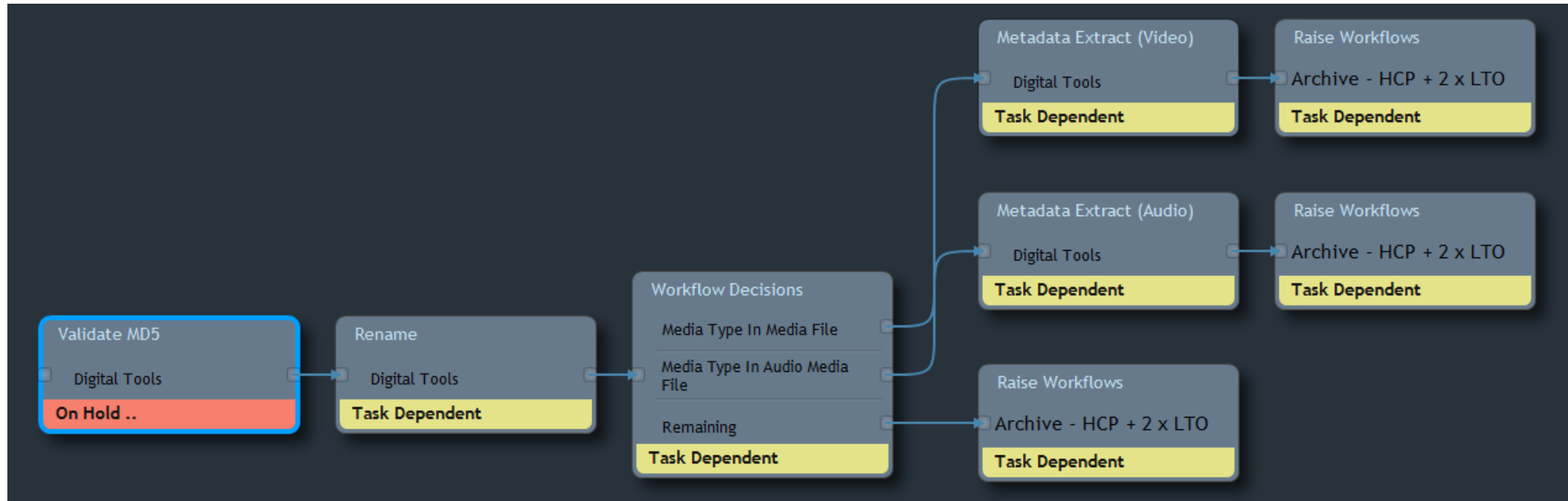
Carriers **Default** > ↔

Barcode ID	Control Symbol	Format	VS	Titles	Versions	Items	# Tit...	Status	Location	Last Location
8415232	32/1/300 Volume 3	CA2 - 16mm film reel	1.5	May Day Melbourne 1953	Primary Version	1 of 1	1	IN	SYD WV 78 7 2 7 / Container#B1745236	SYD WV 78 7 2 7
B1745236		ACO1 - 16mm film (4...		NAA Container		1	0	IN	SYD WV 78 7 2 7	SYD WV 78 7 2 7

# Mediaflex – Workflows



# Mediaflex – Workflows



# Born-Digital File Format Metadata

Filename	Format	Video_Forma...	Video_Forma...	Audio_Format_List	Audio_Codec...	Format	Format_Com...	Format_Profile	General_Inte...	Video_Intern...	Audio_Intern...
C5725_210890_086145.mov	Apple QuickTime Movie	AVC	AVC	AAC	AAC LC	MPEG-4	MPEG-4	QuickTime	video/mp4	video/H264	
C5647_NN1613S042S00MA1_145047.mov	Apple QuickTime Movie	JPEG	JPEG	PCM	PCM	QuickTime	QuickTime		video/quicktime	video/JPEG	
A14412_20_057499.mov	Apple QuickTime Movie	ProRes	ProRes	AAC	AAC LC	MPEG-4	MPEG-4	QuickTime	video/mp4		
C5647_905010_018089.mov	Apple QuickTime Movie	DV	DV	PCM	PCM	MPEG-4	MPEG-4 DVCP...	QuickTime	video/mp4	video/DV	
C5647_NC1571H179S00MA1_047321.mov	Apple QuickTime Movie	MPEG Video	MPEG Video	PCM / PCM	PCM / PCM	MPEG-4	MPEG-4 IMX 50	QuickTime	video/mp4	video/MPV	
C5647_NU1605H117S00MA1_064503.mov	Apple QuickTime Movie	DV	DV	PCM / PCM / PCM / P...	PCM / PCM / ...	MPEG-4	MPEG-4	QuickTime	video/mp4	video/DV	
C5647_NU1835S036S00MA1_216079.mov	Apple QuickTime Movie	JPEG	JPEG	PCM	PCM	MPEG-4	MPEG-4	QuickTime	video/mp4	video/JPEG	
C5647_NN1713S034S00MA1_329548.mov	Apple QuickTime Movie	DV	DV	PCM	PCM	QuickTime	QuickTime		video/quicktime	video/DV	
C5647_NU1836D034S00MA1_374854.mov	Apple QuickTime Movie	DV	DV (DVCPRO)	PCM	PCM	QuickTime	QuickTime DV...		video/quicktime	video/DV	
C5700_1504357_046606.mp4	MPEG4	AVC	AVC	AAC	AAC LC	MPEG-4	MPEG-4	Base Media	video/mp4	video/H264	
D5593_2005-WAV01_147047.mpg	MPEG-2 Video	MPEG Video	MPEG Video	MPEG Audio	MPEG-1 Audio...	MPEG-PS	MPEG-PS		video/mpeg	video/MPV	audio/mpeg
C5725_516911_102797.mxf	JPEG 2000 MXF	JPEG 2000	JPEG 2000	PCM	PCM	MXF	MXF	OP-1a	application/mxf	image/jp2	
C5647_NC1801H163S00MA1_192094.mxf	MPEG-2 MXF	VC-3	VC-3	PCM / PCM / PCM / P...	PCM / PCM / ...	MXF	MXF	OP-1a	application/mxf		