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


• 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.
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
Metadata

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

• Mediaflex uses MedialInfo 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 MedialInfo metadata with metadata from the Samma xml side-car file.
• All other video has a general video schema containing a subset of key MedialInfo fields.
• Agency transfers has a MedialInfo schema that captures all metadata that MedialInfo is able to capture.
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.
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:
  o BWF for audio
  o JPEG2000 in an MXF wrapper for video
  o 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
# Mediaflex – Title Search

<table>
<thead>
<tr>
<th>Title</th>
<th>Format</th>
<th>Filename</th>
<th>Size</th>
<th>Checksum</th>
<th>Status</th>
<th>Location</th>
<th>Last Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>May Day Melbourne 1953</td>
<td>Apple QuickTime Movie</td>
<td>A8703_32-1-300 Volume 3_508906.mov</td>
<td>33.3 GB</td>
<td>25BB1E5D9C831BB99FAD2DDFB9AE25DE</td>
<td>TMD_ARCHIVE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>May Day Melbourne 1953</td>
<td>Apple QuickTime Movie</td>
<td>A8703_32-1-300 Volume 3_508906.mov</td>
<td>33.3 GB</td>
<td>25BB1E5D9C831BB99FAD2DDFB9AE25DE</td>
<td>TMD_ARCHIVE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>May Day Melbourne 1953</td>
<td>Apple QuickTime Movie</td>
<td>A8703_32-1-300 Volume 3_508906.mov</td>
<td>33.3 GB</td>
<td>25BB1E5D9C831BB99FAD2DDFB9AE25DE</td>
<td>TMD_ARCHIVE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>May Day Melbourne 1953</td>
<td>MPEG4</td>
<td>A8703_32-1-300 Volume 3_508906.mp4</td>
<td>153.9 MB</td>
<td>F8DD628592606D71BF6453B81E0109</td>
<td>ACCESS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>May Day Melbourne 1953</td>
<td>MPEG4</td>
<td>A8703_32-1-300 Volume 3_508906.mp4</td>
<td>153.9 MB</td>
<td>F8DD628592606D71BF6453B81E0109</td>
<td>ACCESS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>May Day Melbourne 1953</td>
<td>16_N.B&amp;W_D.Aw.A</td>
<td>A8703_32-1-300 Volume 3_508906.mp4</td>
<td>1 MB</td>
<td>ff4a7601b0a182b99a480ad55b55ab928</td>
<td>ACCESS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>May Day Melbourne 1953</td>
<td>ST_DBC_MF</td>
<td>A8703_32-1-300 Volume 3_508906.mp4</td>
<td>1 MB</td>
<td>ff4a7601b0a182b99a480ad55b55ab928</td>
<td>ACCESS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>May Day Melbourne 1953</td>
<td>JPEG 2000 MXF [ARC]</td>
<td>A8703_32-1-300 Volume 3_508906.mp4</td>
<td>1 MB</td>
<td>ff4a7601b0a182b99a480ad55b55ab928</td>
<td>ACCESS</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**National Archives of Australia**
Mediaflex – Workflows
Mediaflex – Workflows
## Born-Digital File Format Metadata

<table>
<thead>
<tr>
<th>Filename</th>
<th>Media</th>
<th>Format</th>
<th>Audio Format</th>
<th>Audio Codec</th>
<th>Format</th>
<th>Format Comp.</th>
<th>Format Profile</th>
<th>Video Int.</th>
<th>Audio Int.</th>
<th>Video Int.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C5723_E10190_086145.mov</td>
<td>Apple QuickTime Movie</td>
<td>AVC</td>
<td>AAC</td>
<td>AAC LC</td>
<td>MPEG-4</td>
<td>MPEG-4</td>
<td>QuickTime</td>
<td>video/mp4</td>
<td>video/mp4</td>
<td>video/H264</td>
</tr>
<tr>
<td>C6447_NN16135042500MA1_145047.mov</td>
<td>Apple QuickTime Movie</td>
<td>JPEG</td>
<td>PCM</td>
<td>PCM</td>
<td>MPEG-4</td>
<td>MPEG-4</td>
<td>QuickTime</td>
<td>video/quicktime</td>
<td>video/JPEG</td>
<td></td>
</tr>
<tr>
<td>A14412_20_057499.mov</td>
<td>Apple QuickTime Movie</td>
<td>ProRes</td>
<td>AAC</td>
<td>AAC LC</td>
<td>MPEG-4</td>
<td>MPEG-4</td>
<td>QuickTime</td>
<td>video/mp4</td>
<td>video/mp4</td>
<td>video/MPV</td>
</tr>
<tr>
<td>C5647_905010_018089.mov</td>
<td>Apple QuickTime Movie</td>
<td>DV</td>
<td>PCM</td>
<td>PCM</td>
<td>MPEG-4</td>
<td>MPEG-4</td>
<td>QuickTime</td>
<td>video/mp4</td>
<td>video/DV</td>
<td>video/MPV</td>
</tr>
<tr>
<td>C6447_NCI571H77950MA1_047321.mov</td>
<td>Apple QuickTime Movie</td>
<td>MPEG Video</td>
<td>MPEG Video</td>
<td>PCM / PCM</td>
<td>MPEG-4</td>
<td>MPEG-4</td>
<td>QuickTime</td>
<td>video/mp4</td>
<td>video/mp4</td>
<td>video/MPV</td>
</tr>
<tr>
<td>C5647_NUI6650H177500MA1_064503.mov</td>
<td>Apple QuickTime Movie</td>
<td>DV</td>
<td>PCM / PCM / PCM / R...</td>
<td>PCM / PCM / ...</td>
<td>MPEG-4</td>
<td>MPEG-4</td>
<td>QuickTime</td>
<td>video/mp4</td>
<td>video/DV</td>
<td>video/DV</td>
</tr>
<tr>
<td>C5647_NUI663503650MA1_216079.mov</td>
<td>Apple QuickTime Movie</td>
<td>JPEG</td>
<td>PCM</td>
<td>PCM</td>
<td>MPEG-4</td>
<td>MPEG-4</td>
<td>QuickTime</td>
<td>video/mp4</td>
<td>video/mp4</td>
<td>video/PEG</td>
</tr>
<tr>
<td>C5647_NN171503450MA1_249548.mov</td>
<td>Apple QuickTime Movie</td>
<td>DV</td>
<td>PCM</td>
<td>PCM</td>
<td>QuickTime</td>
<td>QuickTime</td>
<td>video/quicktime</td>
<td>video/DV</td>
<td>video/PEG</td>
<td></td>
</tr>
<tr>
<td>C5700_1504357_041606.mp4</td>
<td>MPEG4</td>
<td>AVC</td>
<td>AAC</td>
<td>PCM</td>
<td>MPEG-4</td>
<td>MPEG-4</td>
<td>QuickTime</td>
<td>video/mp4</td>
<td>video/mp4</td>
<td>video/MPV</td>
</tr>
<tr>
<td>D5993_2005-WAV01_147047.mp4</td>
<td>MPEG-2 Video</td>
<td>MPEG Video</td>
<td>MPEG Audio</td>
<td>MPEG-1 Audio...</td>
<td>MPEG-PS</td>
<td>MPEG-PS</td>
<td>video/mpeg</td>
<td>video/MPV</td>
<td>audio/mpeg</td>
<td></td>
</tr>
<tr>
<td>C5725_316911_102797.mp4</td>
<td>JPEG 2000 MXF</td>
<td>JPEG 2000</td>
<td>PCM</td>
<td>PCM</td>
<td>MXF</td>
<td>MXF</td>
<td>OP-1a</td>
<td>application/mxf</td>
<td>image/jp2</td>
<td></td>
</tr>
<tr>
<td>C5647_NCI801H16350MA1_192094.mp4</td>
<td>MPEG-2 MXF</td>
<td>VC-3</td>
<td>VC-3</td>
<td>PCM / PCM / PCM / R...</td>
<td>PCM / PCM / ...</td>
<td>MXF</td>
<td>MXF</td>
<td>OP-1a</td>
<td>application/mxf</td>
<td>image/jp2</td>
</tr>
</tbody>
</table>