

Audiovisual Digitisation: Video Ingest Workflows



Harkiran Dhindsa, Audiovisual Digitisation Mangager, <u>H.Dhindsa@wellcome.org</u> Ashley Ray, Digital Workflow Specialist, <u>A.Ray@wellcome.org</u>

April 2021

Wellcome Collection

A free museum and library exploring health, life and our place in the world.





Film, video & audio collections

- Thousands of audiovisual items held in our AV stores and archive collections
- Recordings cover diverse aspects of healthcare and medicine through 20th century and beyond
- Many non-professional productions, created by people personally connected to the subject matter and are unique to Wellcome Collection
 - it is especially important for us to preserve this material.



Background to AV Digitisation at Wellcome

- Digitisation of selected film & video titles since 2009
- Over 1000 moving image titles digitised & ingested for preservation
- Over 850 videos available to view online at wellcomecollection.org
- Responded to ad-hoc digitisation requests on a small scale.



Preservation of at-risk formats

- Magnetic tape at risk of degrading!
 Video tapes especially at risk
- Film reels also vulnerable to degrading
- Playback equipment becoming obsolete

Move to large-scale digitisation

- Main aim: strategic preservation
 - to complete digitisation of priority at-risk film and video titles within 5 years.
- Digitisation is carried out offsite by specialist AV digitisation contractor
 - deliveries sent every 2 months.



Developing digitisation workflows for video ingest

Workflow considerations:

Preservation

- Lossless format required to meet standards for long-term archiving
- High-resolution capture
 - Avoid re-handling of source material; do not want to revisit material for re-digitisation
 - 2K scans of film

Access

- Preservation master files are of enormous sizes!
- Require compressed files in user-friendly formats

Delivery

• Ease of delivery of preservation/access files to users.



Setting the scene



- Goobi
- Wellcome Storage Service
- DDS/DLCS



New video file type

- Container: MXF
- Codec: JPEG 2000



New tools

Amazon Web Services tools



New Cloud-based terms & tools

S3

Amazon Web Service (AWS) Simple Cloud Storage

Bucket

where S3 stores files, like a file folder

Lambda

a piece of code written in AWS that does a certain thing when an event occurs in a bucket

MediaConvert

a powerful AWS tool that can convert media file types

Glacier Deep Store

S3 cold storage; a place in the Cloud to store large, less frequently accessed files cheaply



Film Pre-Ingest Workflow



bucket

W

Film Ingest Workflow





Goobi's internal film ingest workflow

VMXF_MP4_JPG_PDF_or_MPG_JPG_PDF

🗮 Details				
Number	Title	Status		
10	8 Bibliographic import			
20	8 Video data import			
25	Q Move files			
40	8 Edit METS			
100	📽 Export with PREMIS data			
105	🎕 automatic MMO archive status check			
110	🍪 bagit creation and upload			
115	Send ingest request to archive service			
120	📽 DDS API Call			
140	📽 15 days delay			
150	📽 Image removal			

METS Decisions



bnumber.xml Anchor file



bnumber_0001.xml Film

bnumber_0002.xml Transcript

One METS anchor file, several individual METS

 Anchor METS file ties together the METS for individual parts of ingest



bnumber.xml

New METS Option 1: One METS file with MIMETYPES only

- Film, Transcript, Poster Image in 1 file group, files described by MIMETYPES
- DDS action decided by MIMETYPE in hard-coded rule



bnumber.xml

New METS Option 2: One METS file with File Groups and Usage Attributes

- Film, Transcript, Poster Image in multiple file groups with usage attributes
- Usage attributes will direct DDS

Future proofing: Deciding on METS

<pre><mets:filegrp id="OBJECTS" use="ACCESS"> <mets:file admid="AMD_0001" id="FILE_0001_OBJECTS" mimetype="video/mp4"> <mets:flocat loctype="URL" xlink:href="objects/b30496160_0002.mp4"></mets:flocat> </mets:file> </mets:filegrp> </pre>	ID: Objects, Use: Access This file is served up for online access
<pre><mets:filegrp id="POSTER IMAGE" use="POSTER"> <mets:file admid="AMD_0003" id="FILE_0001_POSTER" mimetipe="Image/jpeg"></mets:file></mets:filegrp></pre>	ID: Poster Image, Use: Poster This image file is applied to a film or audio file player to display before play is pressed
<pre><mets:filegrp id="MASTERS" use="PRESERVATION"> <mets:file admid="AMD_0002" id="FILE_0001_MASTERS" mimetype="application/mxf"> <mets:flocat loctype="URL" xlink:href="objects/b30496160_0003.mxf"></mets:flocat> </mets:file> </mets:filegrp> </pre>	ID: Masters, Use: Preservation This file is for preservation only and is put in Glacier deep store
<pre><mets:filegrp id="TRANSCRIPT" use="TRANSCRIPT"> <mets:file ID="FILE_0001_TRANSCRIPT" MIMETYPE="application/pdf" ADMID="AMD_0004</td><td>ID: Transcript, Use: Transcript This file is served online as the accompanying transcript to a film or audio file</td></mets:filegrp></pre>	ID: Transcript, Use: Transcript This file is served online as the accompanying transcript to a film or audio file

Goobi's internal film ingest workflow

VMXF_MP4_JPG_PDF_or_MPG_JPG_PDF

Details				
Number	Title	Status		
10	8 Bibliographic import			
20	8 Video data import			
25	08 Move files			
40	8 Edit METS			
100	🕸 Export with PREMIS data			
105	🚓 automatic MMO archive status check			
110	🕫 bagit creation and upload			
115	Send ingest request to archive service			
120	© DDS API Call			
140	🚓 15 days delay			
150	📽 Image removal			

Film Ingest Workflow







BMA Hastings Festival.

Date 1951

Videos Online

Available online



License Attribution 4.0 International (CC BY 4.0)

Looking back on the process and new workflow

Things that could be improved

- To take advantage of our new Cloud platform, we had to wait for our new Cloud platform
- We had to leave a hefty backlog lingering
- Learning curve in using AWS
- Only select members of the team can do the ingest (permissions to the Cloud are locked down)
- Need to rely on separate team (Platform dev team) to help us make changes or troubleshoot on AWS

Things that went well

- Achieved our workflow goals
 - Files are safely stored before ingest
 - No downloading/uploading necessary
 - Can convert automatically and on the fly
 - Poster image and transcript ingested with master files
- Had support of developers
- Learned about METS; made good choices for the future
- Documentation done as we went along

More information

- Policies and plans <u>https://wellcomecollection.org/pages/Wvmu3yAAAIUQ4C7F</u>
 - Digitisation Strategy [PDF]
- Library catalogue https://wellcomecollection.org/works
- Digital Engagement blog https://stacks.wellcomecollection.org/
- Developer information https://developers.wellcomecollection.org/
- Github site https://github.com/wellcomecollection
- Public roadmap <u>https://roadmap.wellcomecollection.org/tabs/1-planned</u>