

Challenges and solutions in broadcast archives

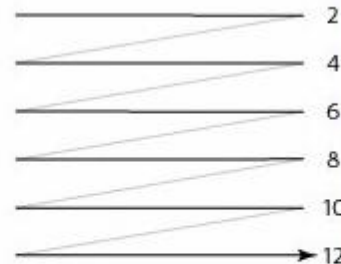
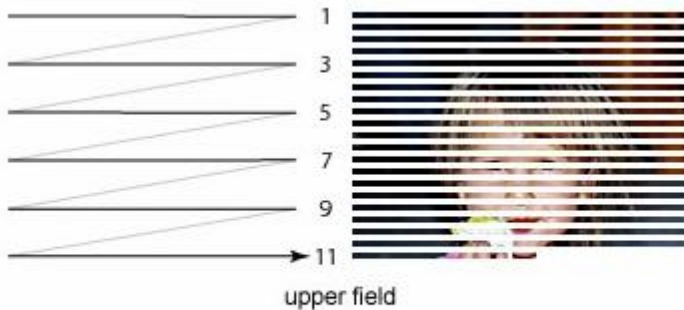
BBC Research and Development
John Zubrzycki

Contents

1. Broadcast TV and Radio
2. BBC's A/V archive and tapeless digital production
3. Digital formats for A/V preservation
4. Storage of digital A/V content

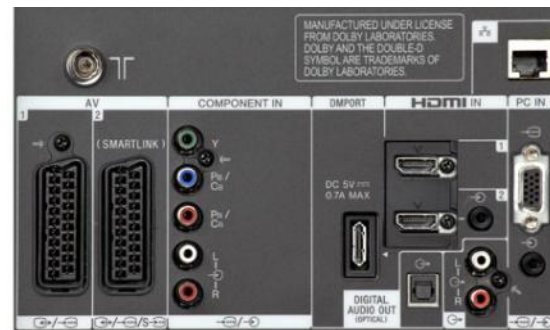
1. Broadcast TV and Radio

- Standard definition:
 - 720 pixels x 576 lines
 - 4x3 or 16x9 aspect ratio
 - 25 fps (Europe), 30 fps (North America)
 - Interlaced



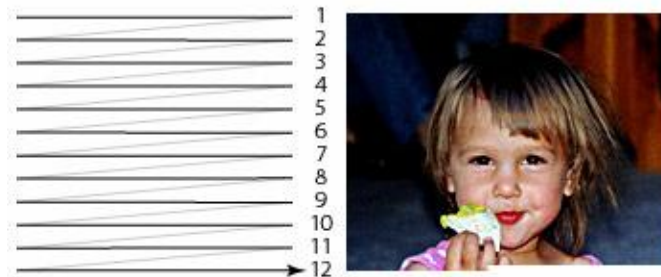
lower field

- Standard definition:
 - Analogue:
 - PAL (Europe), NTSC (North America)
 - Video interfaces:
 - Composite (1 connector)
 - Component RGB or YUV (3 connectors)
 - S-Video (multiway connector)
 - SCART (multitway connector inc. audio)
 - Digital:
 - DVB carrier
 - MPEG 2 video coding
 - Video interfaces:
 - Analogue interfaces often used
 - HDMI (consumer)
 - SDI (professional)



TV formats

- High definition:
 - 16x9 aspect ratio
 - 720p
 - 1280 pixels x 720 lines
 - 50 fps (Europe), 60 fps (North America)
 - Progressive
 - 1080i
 - 1980 (or 1440 pixels x 1080 lines
 - 25 fps (Europe), 30 fps (North America)
 - Interlaced
 - 1080p
 - 1980 pixels x 1080 lines
 - 50 fps (Europe), 60 fps (North America)
 - Progressive



- High Definition:
 - Analogue / hybrid:
 - MUSE (Japan), [HD-MAC (Europe)]
 - Interfaces:
 - Component RGB or YUV (3 connectors)
 - Digital
 - DVB (Europe, satellite and cable), ATSC (North America)
 - MPEG 2 or AVC (aka MPEG2) video coding
 - Video interfaces:
 - HDMI + HDCP (consumer)
 - HD-SDI (professional)
 - Analogue interfaces sometimes used
- Blu-ray disc: HDTV at 25 or 30 fps HD films at 24 fps



- 3D-TV
 - Side-by-side format
 - Left and right eye pictures placed within an HD frame



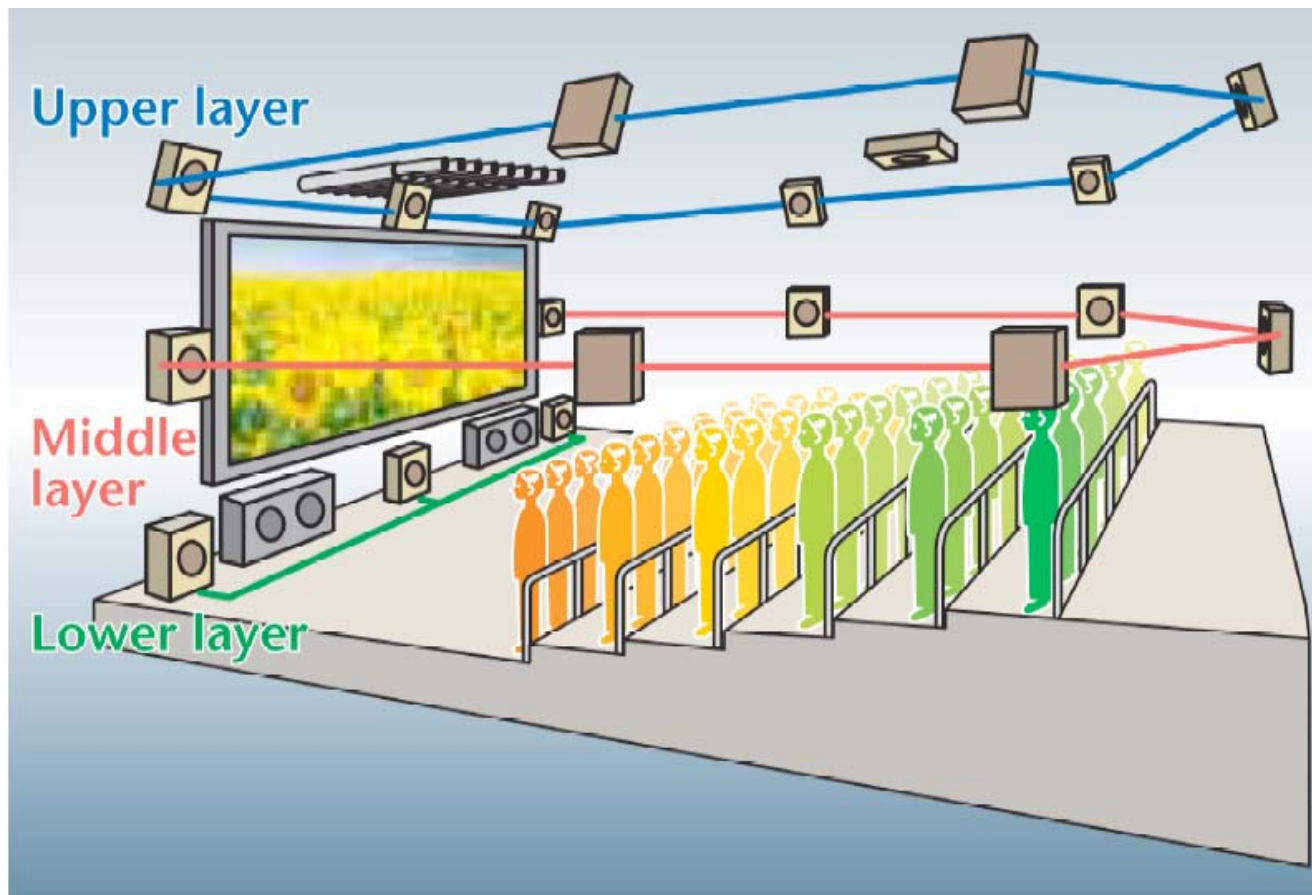
- Ultra-HD
 - 4k
 - 4096 pixels x 2160 lines
 - 24 fps
 - 16:9 container
 - Based on SMPTE Digital Cinema Initiative
 - Super Hi-Vision (SHV)
 - SHV1: 3840 pixels x 2160 lines (4k)
 - SHV2: 7680 pixels x 4320 lines (8k)
 - 50, 60 or 120 fps
 - 16:9 aspect ratio
 - Based on HDTV (ITU Rec.709)



- Radio:
 - AM (MW, LW, SW), FM (Analogue)
 - Mono or stereo (FM)
 - DAB (Europe), HD Radio (North America) (Digital)
 - Stereo, mono or multichannel
 - MPEG-1 Layer 2 coding
 - DRM (digital shortwave)
 - Stereo, mono or multichannel
 - MPEG-4 HE AAC v2 coding
- TV:
 - Analogue FM
 - Mono
 - NICAM (digital)
 - Stereo or multilingual
 - DVB (digital)
 - Mono, stereo or 5.1, Audio Description
 - MPEG1 Layer 2, MPEG-2 AAC-LC (for HDTV)



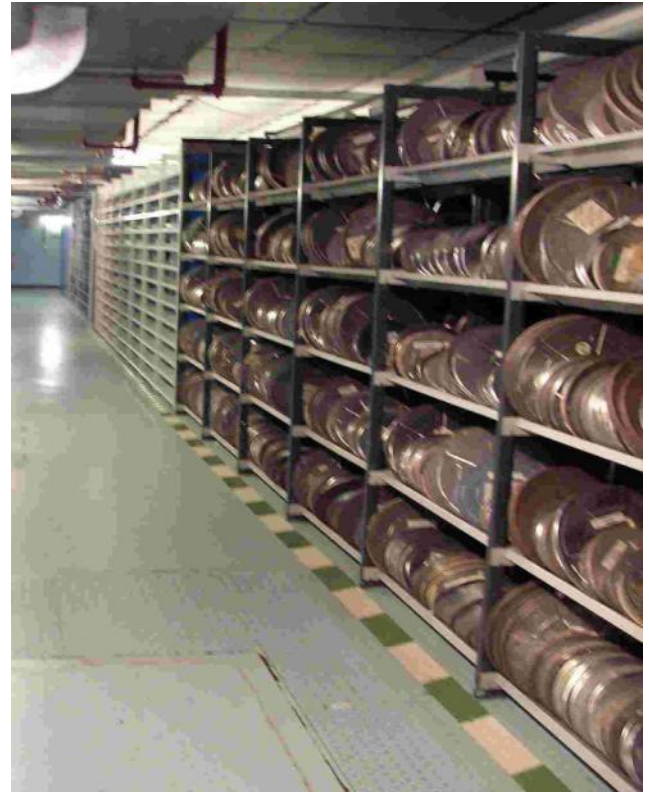
22.2 channel audio for Super Hi-Vision TV



2. BBC's A/V archive and tapeless digital production

1. BBC Archives

- 650k hours video
- 350k hours audio
- 2M stills; ¼ online
- 3M items sheet music
- 400k “pronunciations”
- 1.5M titles in “grams library”
- BBC business documents
- 100 km of shelves.



1. BBC Archives

- 95% Internal use by BBC
- Established workflow based on tapes and film
- Browse or broadcast quality tapes can be supplied to programme-makers in 24 hours
- Infax database available to BBC staff to find content
- LonClass classification system for powerful relational search
- Experienced Media Managers can help programme makers find suitable content and obtain rights clearances.

1. BBC Archives

- Moved to new purpose-built premises in Perivale, west London, this year
- Migrating to digital file-based operations and being integrated into the BBC's digital production system: Fabric
- D3 digital videotape was used from the early 1990's:
 - Migrated 2" and 1" analogue videotape to D3
 - But D3 is not file-based
- D3 preservation project migrating D3 to digital files:
 - PAL transform decoder developed by BBC R&D
 - 200 hrs / week
 - Stored as uncompressed files (90 GB/hr) in an MXF wrapper
 - 1 PetaByte / yr
 - Expanding to other tape formats (Digibeta and HD)

- Digitisation of tape using:
 - Commodity IT hardware:
 - Plus video capture card
 - Open source software:
 - <http://ingex.sourceforge.net/index.html>
 - SD codecs:
 - JPEG
 - DVCPRO50 (50Mbps) and DV (25Mbps)
 - IMX 50/40/30 (50/40/30Mbps)
 - Uncompressed 8 bits-per-sample
 - HD codecs:
 - VC-3 at 120Mbps and 185Mbps
 - DVCPRO HD
 - Uncompressed 8 bits-per-sample



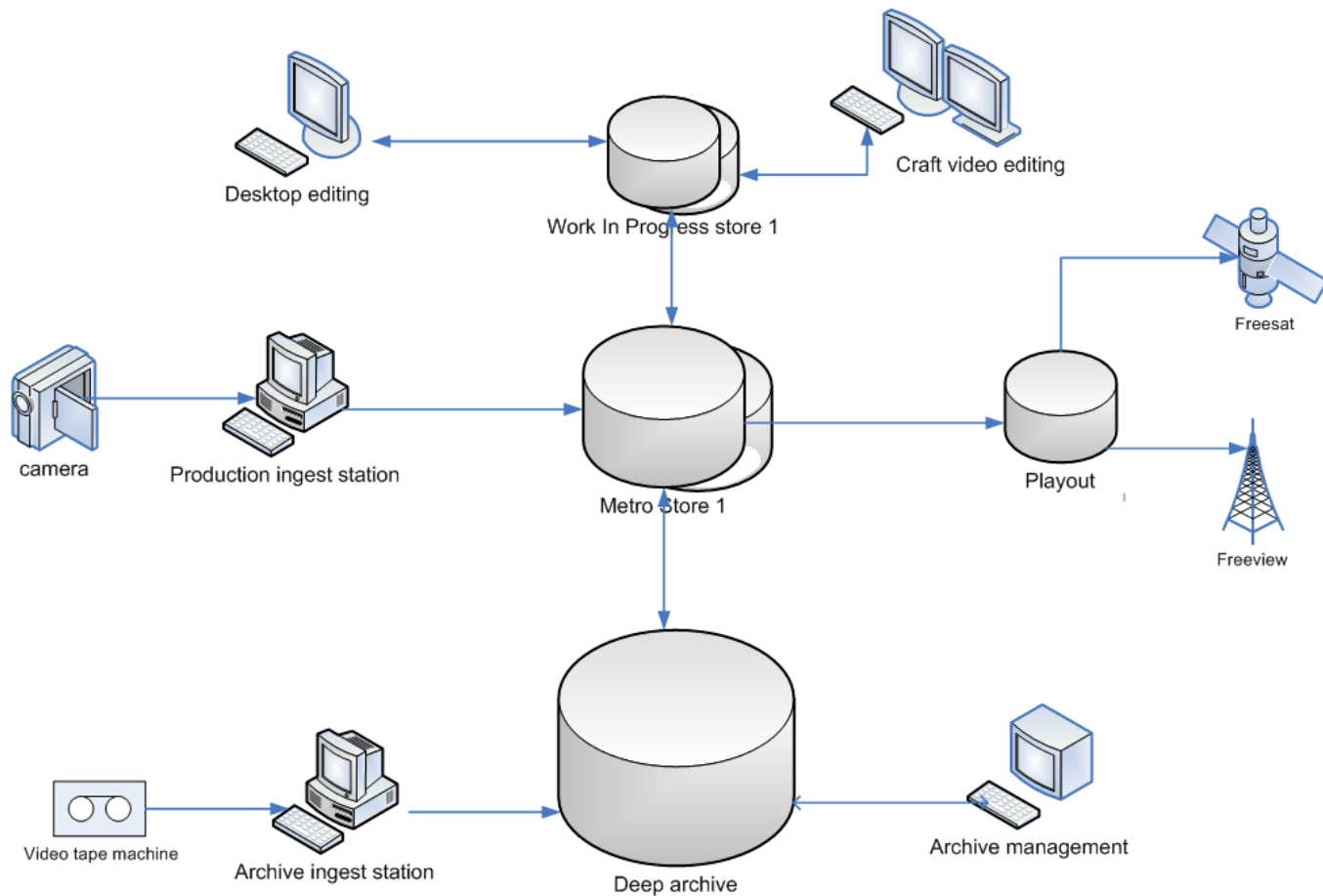
- Restoration and preservation of deteriorating original content:
 - Increases 'value' of archived content
 - In the 1960's video tape was expensive – was widely re-used with loss of key original material:
 - Restoring colour video from old black & white tele-recordings via detection and decoding of colour sub-carrier dot-pattern.



Tapeless production

- Videotape is fragile, expensive and can usually only be transferred for editing etc in real-time.
- BBC Digital Media Initiative project developing the Fabric desktop production system:
 - Ingest from camera solid state memory card faster than real-time
 - Addition of metadata about the production only once
 - Automatic inclusion of metadata from cameras etc.
 - Programme-makers rough-cut edits at desktop via browser
 - Working with browse quality proxies to low network load
 - Access to archive browse store from desktop

Tapeless production



- Edit decisions sent to craft post-production processes for making final programme
- Broadcast quality content sent to ‘craft’ automatically from Work in Progress store, Metro store or deep archive
- Final programme, metadata and selected rushes sent to deep archive for preservation

3. Digital formats for preservation

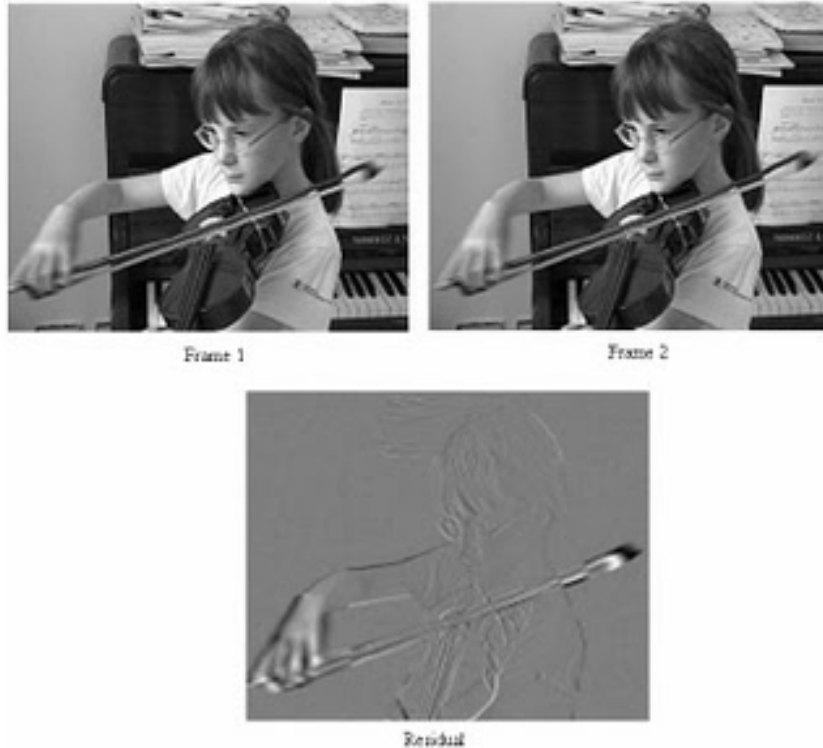
A/V preservation

- Archives tend to work of the principle that they preserve what they are given.
- Works well for film where the medium has stayed stable for many 10's of years
- Video tape is less stable, also formats become obsolete after several years requiring dubbing of content onto a new format
- Digital files separate the format of the content from the parameters of the physical medium, easing transfer to a new medium as the old one become obsolete



A/V preservation

- Video production formats often use compression (video encoding) to save space in the camera memory



Types of compression

- Production / delivery formats:
 - MPEG1, MPEG2, MPEG4, DVC family, IMX, AVC-I
 - Typically 10 – 40:1 compression ratio on file size
 - Acceptable if content created in that format originally
- Light (mezzanine) compression
 - SMPTE VC2 (Dirac), VC3 (DNx)
 - Apple ProRes
 - Typically 4 – 6:1 compression ratio
 - Visually lossless
- Lossless compression:
 - JPEG2000 lossless
 - Typically 2:1 compression ratio
 - Can get original bits back
- Uncompressed:
 - Preserves original bits
 - Uses most storage space (expensive)



- Video coding reduces storage requirements
- Each compression technique adds its own type of impairments
- Coding concatenation builds up impairments:
 - Particularly if compression types are mixed
 - ‘Last Straw’ effect result in sudden appearance of impairments
- Compressed content is more fragile in the presence of errors
- Long-term sustainability of the decoders
 - Will decoders still work on new operating systems?
 - Is emulation a safe way to run decoders?.
 - Best chance for ongoing support is if all archives use same codecs
- Audio the debate is over:
 - Uncompressed
 - 48 kHz sampling

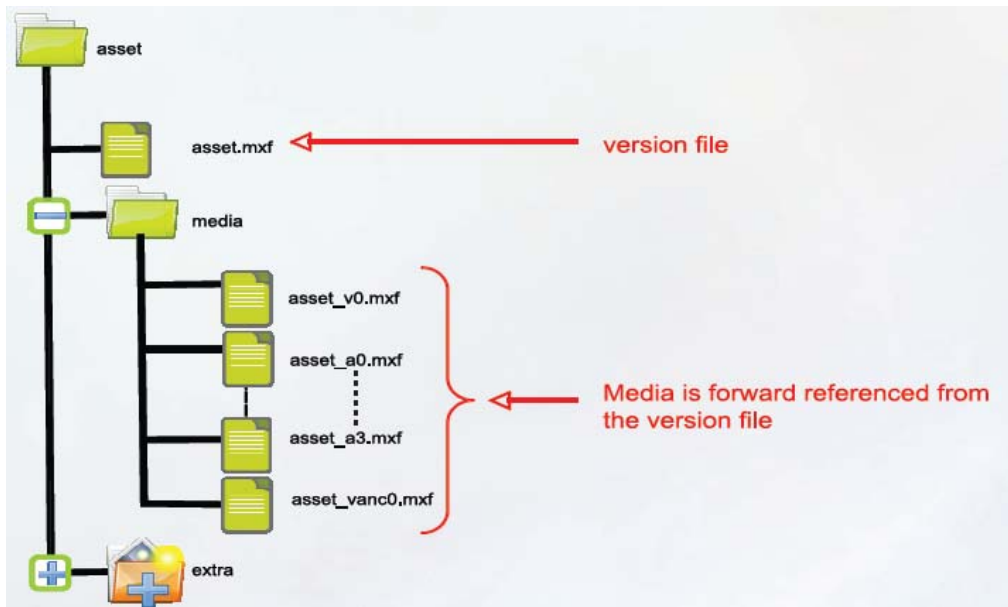
Suggested approach to archiving compressed content

- Accept borne digital content in the file format given
- Store and supply content in same format
- When the production format changes, convert content to lightly compressed, lossless compressed or uncompressed.
 - Transcode to new production format to supply when requested
- Ensure that decoders can still be run when upgrading the storage management system h/w or s/w.
 - If problems found, then convert content to uncompressed
- Do not re-code content more than necessary
- Critically view content on a good quality large display to check for visible impairments
- A/V archives need to agree a set of archive quality light compression codecs for both SD and HD for situations where uncompressed is not practical

- Video and audio content needs to be put in a wrapper along with essential metadata to keep it all together
- Two popular wrappers:
 - MXF
 - Quicktime
- BBC chosen MXF
- MXF variants:
 - OP Atom (for production)
 - OP 1a (for archive)

Packages

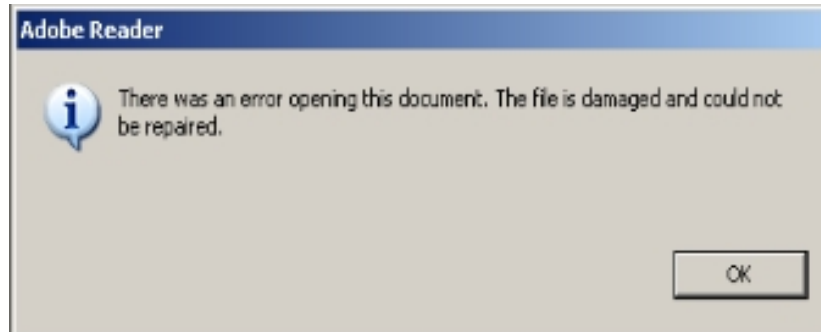
- AS02 package:
 - Contains several MXF files, e.g. different takes
 - Also can contain other files, e.g. scripts
- DPP (Digital Programme Package)
 - Content exchange between organisations
 - Programme in MXF plus other files



4. Storage of digital A/V content

Digital storage

- “We’ve digitised all our content – it’s safe”
- “I just need a place to store the floppy disks”
- Digital technology is evolving fast creating obsolescence in only a few years!
- “Help! I got this message when I tried to open the file”

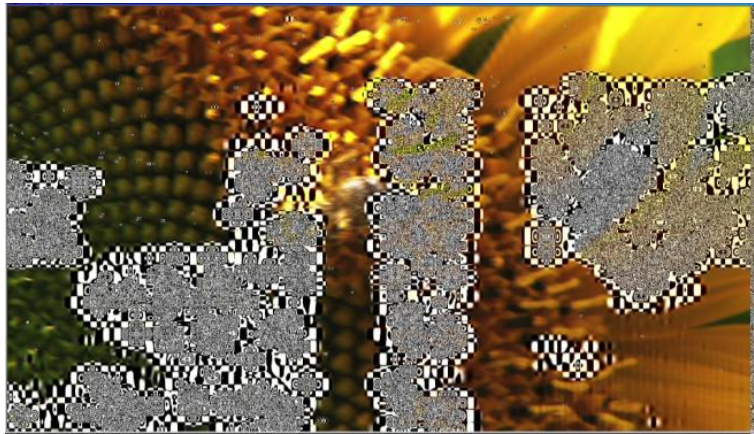


- Preserve content in an error-tolerant way.

BBC digital storage requirements (uncompressed):

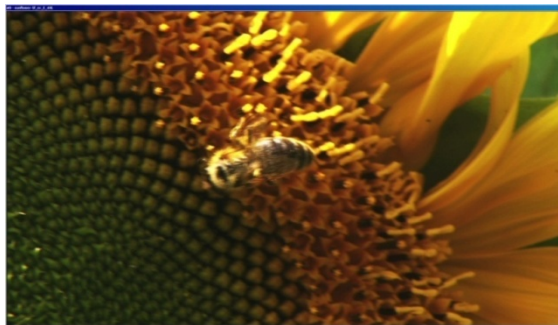
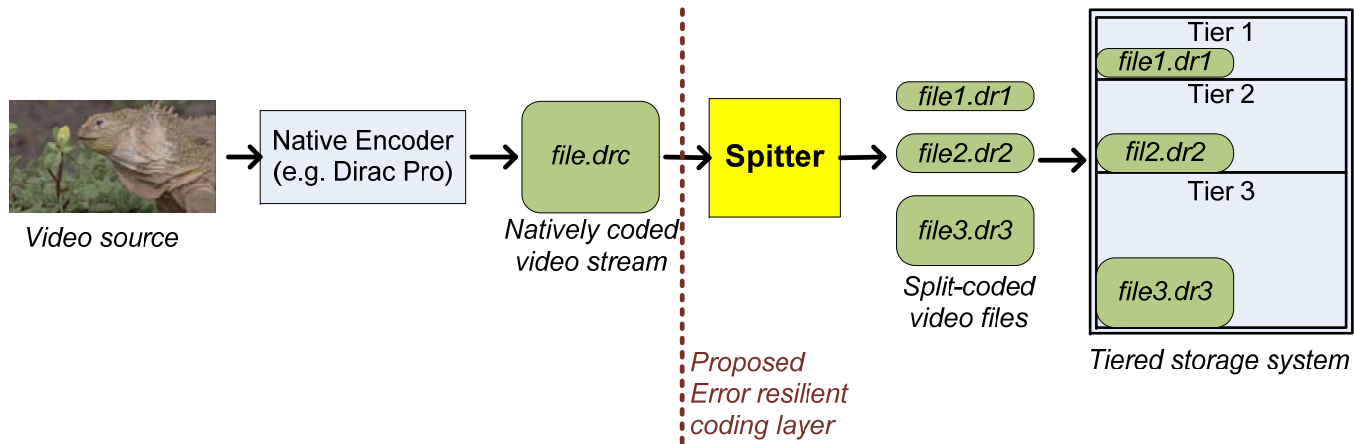
- Preservation:
 - 1500 video items / 800 hours/week = 80 TB per week
 - For 16 years to digitise whole archive
 - 65 PetaBytes
- New content:
 - 300 hours/week = 30 TB per week
 - moving to 120 TB/week with move to HD
 - 100 PetaBytes (in 16 years time)
- Total storage needed in 2026:
 - 165 PetaBytes (one copy only) (Data tape cost £5.8M)
 - 330 PetaBytes (two copies) (£11.6M)

- Digital storage is not perfect:
 - CERN found error rate in storage of 1 in 10^7
 - Equivalent to 33 Gbytes in 330 PetaBytes
 - Errors highly bunched:
 - e.g. due to RAID failure
 - Effect of errors on coded pictures:



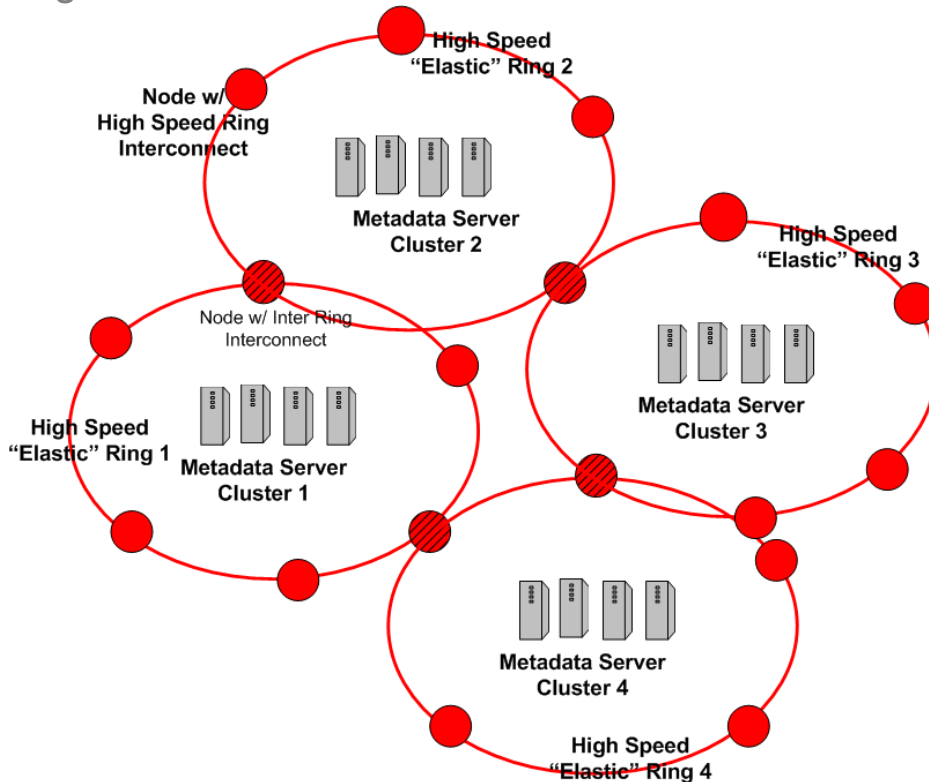
Protection of compressed files

- Avatar-m project developed codec header protection



AV Storage

- Video files can be very large (HD = 400 GB /hr)
- Large storage systems need **scalable** high throughput networks to avoid congestion



- File management needs to be adapted to handling large files:
 - E.g. use of file chunking to allow access to a file while it is being used
- Storage system lifecycle management:
 - Very large Peta-Exabyte storage systems
 - Storage installed over years
 - Might include external cloud storage
 - Sections of storage regularly becoming unserviceable
 - Data migration from old to new storage needs to be integral to storage management system
- Data integrity checking needs to work at the video and audio content level – not just file checksums
- Avatar-m project studied these issues:
 - <http://www.avatar-m.org.uk>



- PrestoPrime project is developing techniques for digital A/V archiving:
 - <http://www.prestoprime.org/>
- PrestoCentre new source of digital A/V archive information:
 - <http://www.prestocentre.eu/>



Conclusions

- BBC Archives will need at least 330 PetaBytes storage by 2016
- A/V formats regularly change, so content needs to be migrated to a long-term archive format
- Technology choices important for reliable long-term storage:
 - Compression choices – avoid concatenating codecs
 - Storage system choices – ensure system scales
- PrestoCentre source of advice for archives
- Archives should work together to present common requirements to industry; e.g to agree:
 - Light compression standards for SD and HDTV

I would like to thank my colleague Richard Wright for his advice in preparing this presentation and for supplying some of the information