Digital Archiving of Audiovisual Media

What's Worse: the Problem or the Solution?

Richard Wright BBC Information & Archives

European Digital Cultural Heritage

Salzburg 2006.06.22

The Problem: Analogue Media

Decaying
Obsolete
Fragile

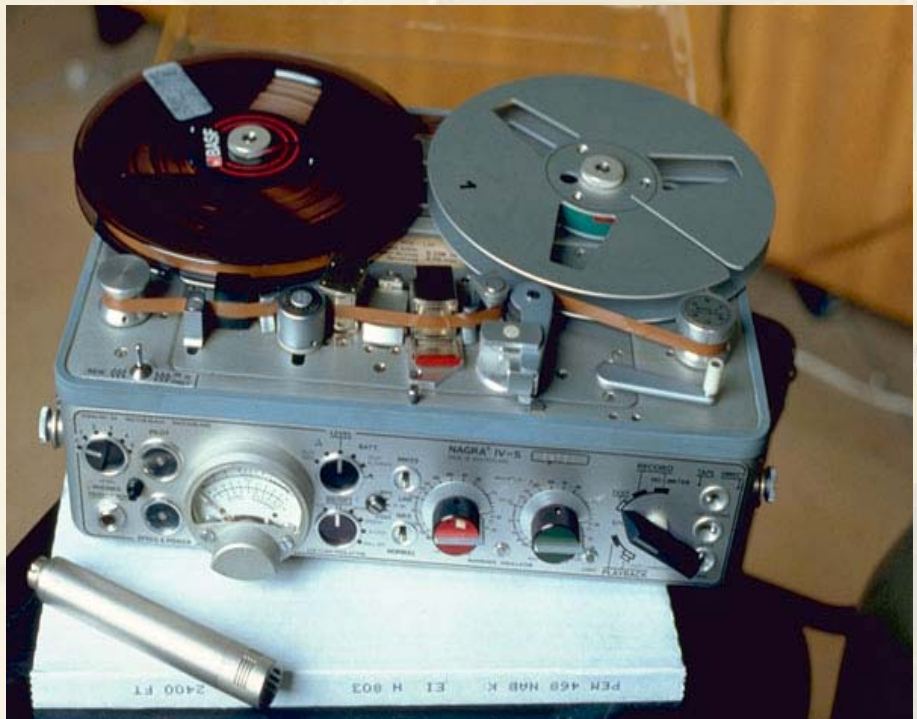
Presto Survey, 2001
5 million hours of holdings

Decaying Obsolete Fragile

- Obsolescence: at least 2/3 of the material
- Deterioration: approximately 1/3 of the material
- Fragile media: roughly 1/4 of the material

Overall: 70% of holdings have problems

The Solution: digitisation



Size of the Problem – in Europe

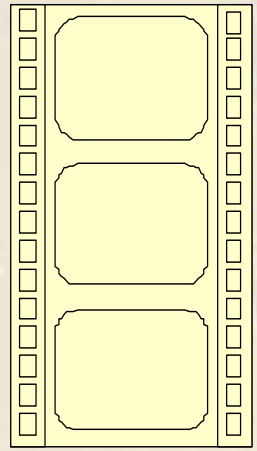
- Presto: found 5 million hours 2001
  - Mainly broadcast archives
- Prestospace: found 10 million hours 2004
  - Broadcast and large national collections
- TAPE: found **additional** 20 million hours
  - In collections not covered previously
- UNESCO estimate: 200 million hours worldwide

Where is the material?

- Broadcast archives 30% (roughly)
- National collections 15%
- Other major collections 15%
- Small and specialist collections 40%

NB: all these figures refer to archived material ONLY

What is the material



- Film:
  - cinema (commercial and heritage)
  - “other”
- Broadcasting:
  - audio, on many formats
  - video, again on many formats
  - and lots of film!
- Heritage, library, archive, commercial and academic collections

What are audiovisual collections good for?

- Documentation, Enrichment, Appeal
- Actuality of the 20th Century

What if we had:

- “Meet the Borgias” on MTV
- Big Brother in a plague city
- Mona Lisa on a talk show
- Footage of Agincourt
- a BBC studio session with Mozart

Full details on what audiovisual media exists, and where – are available from PrestoSpace and TAPE

Coping with the digital solution:

Analogue formats lasted a few decades

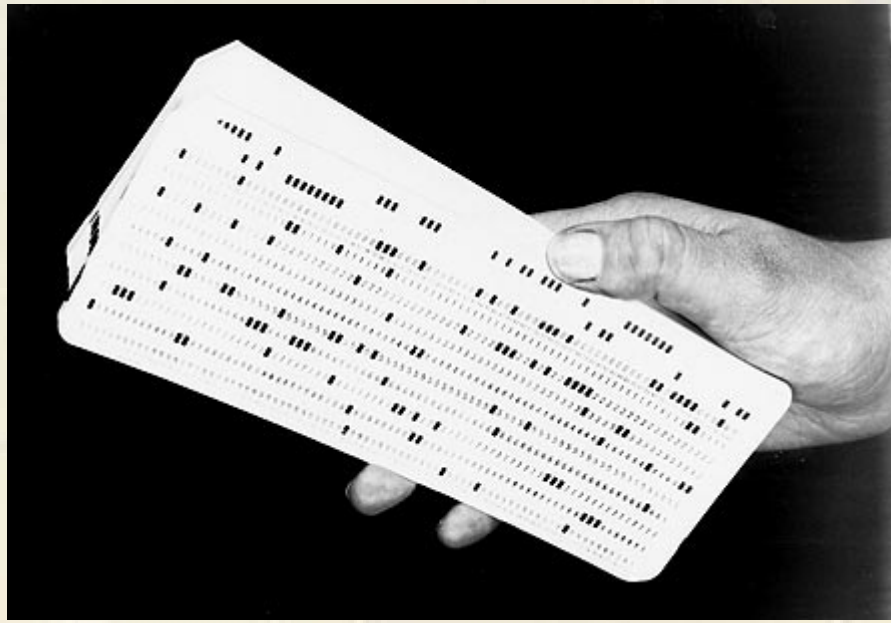
- (except film, which can last a century or more)

- The time between introduction and obsolescence has been shrinking, especially in video formats.
  - The first BBC digital video format had a **ten** year active life.
- For data formats (datatape, hard drives), the life-cycle is even shorter.
  - New datatape formats emerge about every **three** years.

Advice from the BBC and EC project PrestoSpace.

- Choice: which digital technology
- Management: cost-effective and secure

The new problem:
digital media
digital media



- How long does it last?
- How reliable is it?
- How much maintenance does it need?
- And how much does that cost?
- How can we achieve “archive preservation standards”?

Aspects of Digital Management

- Media
- Multiple copies
- Maintenance
- Migration

Media

- **Datatape** is cheaper that hard drives
  - But needs an expensive tape drive
  - And has reliability issues
- **Optical** is cheapest of all
  - But isn't really mass storage (DVD=4.7 GB)
    - New DVD format(s) promise 20 to 100 GB
  - And has reliability issues
- **Hard drives** prices have dropped sharply
  - Easiest to automate management
  - And has reliability issues
- More information from PrestoSpace:
  Prestospace.org ("digitisation & storage")

Multiple copies

- Two copies
  - Two technologies
    - In two places
- But fastest recovery is by mirroring
  - Which means identical technologies
- Big arguments about RAID vs simpler options vs more complex options

Maintenance

- Life cycle management
- Should be every archive's built-in process
- Begins with blank media
  - Then the writing
  - Then the initial checking
  - Then the periodic checking and 'aerobics'
- Ends with migration to the next format

Migration

- A fact of life
- Every five years
- Can involve a lot of manual handling (of datatapes or optical media)
- Or can be nearly transparant (disc upgrades) – but: every three years!
- Needs lossless file formats

Migration Volumes

- **Small sound archive**: 20k hrs mono

  3 hrs/GB => 7 TB = 9 datatapes @ 800 GB
- **Imperial War Museum**: 120 million ft film, plus 6500 hours audio

  Film: 90 ft/min = 5400 ft/hr (35mm, 24 FPS)

  => 22k hrs film; lossless high def = 100 GB/hr
  (which is 300x higher than for audio)

  So: 2.2 petabytes = 2750 datatapes @ 800 GB

  (this is what the BBC call a small migration)

Strategy

- Two copies, two technologies, two locations
- Migration every five years
  - Or continuously!
- “Just in time” purchasing of storage
- Web access: storage costs are zero (if you have lots of access)

BBC Examples

- **6mm audio tape** to audio CD and to BWF files on DVD
  - and Real Audio web-access files
  - Same process for 78 and LP discs
- **U-Matic** to DV-CAM and to MPEG-II files (on DVD)
- **D3** to digibeta and uncompressed onto datatape
  - with MPEG-4 viewing copies

6mm audio tape

- to audio CD
- and to BWF files on DVD
- and Real Audio web-access files

CD / DVD life expectancy > 20 yrs

Expected migration to datatape within 10 years

(Same process for 78 and LP discs)

U-Matic

- to DV-CAM (working format)
- and to MPEG-II files (now held on DVD)

DV-CAM expected to be a production format for < 5 more years

DVD’s expected to last > 20 years

Expected migration of DVDs within 10 years

- To digibeta
- and uncompressed onto datatape
- with MPEG-4 viewing copies

Can you trust mass storage?

- You can't trust anything
- Every options must be a managed option
- Management of:
  - **M**edia
  - **E**quipment
  - **A**ccess
  - **O**bsolescence
  - **D**egradation

Lots of information on the PrestoSpace website, from TAPE training, and from the Preservation Guide wiki

www.prestospace.org

www.knaw.nl/ecpa/tape/

www.bbcarchive.org.uk

Successful Management

Requires

- A plan
  - Covering MEAOD
- A budget
  - BBC now spends +20% on “preservation”
  - Digital maintenance should be much cheaper

A Happy Digital Mass-Storage Future

Fred G Moore

www.prestospace.org

richard.wright@bbc.co.uk