

Sustainable Digital Preservation and Access: A view from public broadcasting

BBC Research and Development
John Zubrzycki

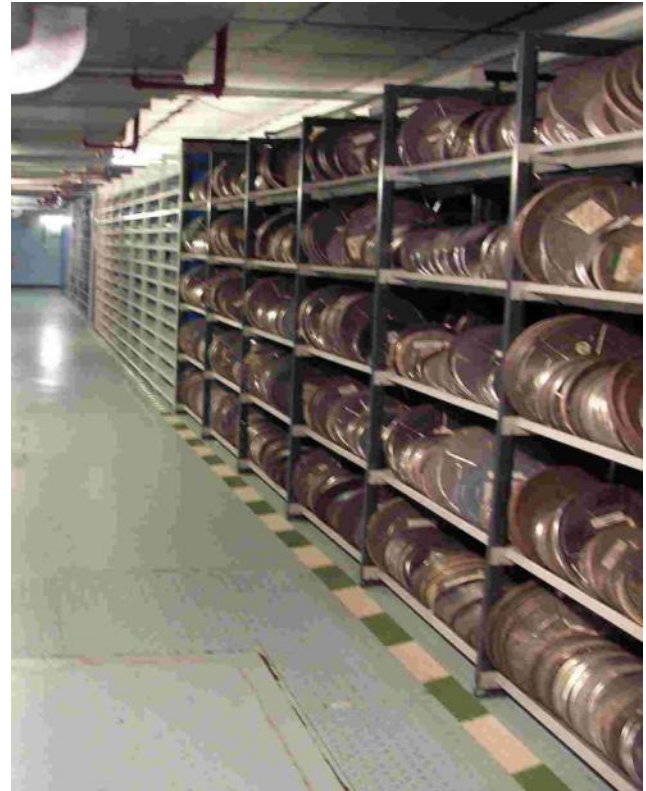
Contents

1. BBC archives
2. Archives at the centre of broadcasting
3. Preservation economics in broadcasting
4. Partnerships

1. BBC Archives

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

- Moving to new purpose-built premises in Perivale, west London, later this year
- Migrating to digital file-based operations
- 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
 - 1 PetaByte / yr
 - Expanding to other tape formats and to increase throughput.

- 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.



2. Archives at the centre of broadcasting

2. Archives at the centre of broadcasting

- BBC charter obligations:
 - To preserve its output
 - To make it available to the public
- Public access is mainly via the BFI and the BL Sound Archive
- Roly Keating appointed BBC Director of Archive Content:
 - Aim: public access via the web by 2020
- What does public web access look like?

2. Archives at the centre of broadcasting: BBC Archive Collections

The screenshot shows the BBC Archive website interface. At the top, there is a navigation bar with the BBC logo, a search bar, and a 'Text only Help' link. Below this is a yellow navigation bar with the following menu items: ARCHIVE, Home, Collections, Subjects, Programmes, People, Meet the Experts, and Help. The main content area features a large featured collection titled 'SWINGOMETER' with a yellow 'Latest Collection' badge. The image shows a man in glasses and a pollster's pendulum chart. To the right, there is a 'Featured Collections' section with three items: 'REMEMBERING CHURCHILL', 'THE CHANGING FACE OF DOCTOR WHO', and 'SECOND WAVE FEMINISM'. Below the main content are three sections: 'Today's Choice' (featuring NEWSNIGHT), 'Around the BBC' (featuring 'IN OUR TIME ARCHIVE'), and 'Elsewhere on the Web' (listing 'NATIONAL ARCHIVES').

BBC Text only Help Search Explore the BBC

ARCHIVE Home Collections Subjects Programmes People Meet the Experts Help

Latest Collection **SWINGOMETER**
The pollster's favourite pendulum

Featured Collections

- REMEMBERING CHURCHILL**
Tributes to a legendary statesman.
- THE CHANGING FACE OF DOCTOR WHO**
How the BBC recast TV's Time Lord.
- SECOND WAVE FEMINISM**
Issues, Attitudes and Opinions.

Today's Choice
Programme of the day
NEWSNIGHT

Around the BBC
IN OUR TIME ARCHIVE
The archive of the programme which explores the history of ideas.

Elsewhere on the Web

- NATIONAL ARCHIVES**
The UK government's official archive.
- REF SCREENLINE**

www.bbc.co.uk/archive

2. Archives at the centre of broadcasting: BBC Archive Collections

The screenshot shows the BBC Archive website interface. At the top, there is a search bar and a navigation menu with options like 'Home', 'Collections', 'Subjects', 'Programmes', 'People', 'Meet the Experts', and 'Help'. The main content area is titled 'ARCHIVE' and features a large banner for the 'Chronicle' collection. The banner includes the text 'About this collection' and a description: 'For 25 years, the BBC's archaeology series took viewers around the world to explore historical excavations and discover long-gone cultures and civilisations. With a mix of live broadcasts and filmed documentaries, 'Chronicle' brought some of the greatest archaeologists of the 20th Century into our homes. In this collection, we look back at a selection of programmes from the series, including reports on ancient Greece and Sutton Hoo, plus a memorable live broadcast from the Silbury Mound.' Below the banner, there are several program thumbnails, each with a title, a broadcast date, and a duration. The thumbnails include 'Roman Goose March' (1966, 10 minutes), 'The Fall of Constantinople' (1967, 33 minutes), 'The Last Days of Minoas' (1967, 48 minutes), 'The Great Iron Ship' (1967, 33 minutes), and 'Cracking the Stone Age Code' (1967, 48 minutes). A 'FIRST BROADCAST 1989' badge is also visible.

This screenshot shows a document scan from the BBC Archive. The document is a letter from George Orwell to Rayner Heppenstall, dated 1946. The text of the letter is visible, starting with 'Dear Rayner,' and discussing the difficulty of writing letters. The scan is presented in a digital format with a search bar and navigation options at the top. A '1946' badge is visible on the right side of the document.

GEORGE ORWELL AT THE BBC Collection

This screenshot shows the 'George Orwell at the BBC' collection page. It features two document scans. The first is a 'MEMO FROM THE BBC'S DIRECTOR OF EMPIRE SERVICES' dated 1941, which recommends George Orwell for employment. The second is a 'LETTER FROM ARTHUR KEITH TO D PEARSON-SMITH ESQ' dated 1941, which is a glowing reference for George Orwell. Both documents are presented in a digital format with a search bar and navigation options at the top. A '1946' badge is visible on the right side of the second document.

Document scanning standards?

2. Archives at the centre of broadcasting

Creative Archive:

- BBC, British Film Institute, Channel 4, Open University...
- Started in 2005
- BBC contributed 500 content items



- BBC pilot ended in 2006
- BFI and Open University pilots still running.



5 Rules of the Creative Archive Licence



No commercial use



Share alike



Give credit



No endorsement



UK only

2. Archives at the centre of broadcasting

- The web is an opportunity to develop new services and applications
- BBC R&D and I&A have been involved in research projects providing web access to content:
 - BIRTH of TV
 - Video Active
 - Spoken Word
- Rights issues either restrict projects to academic access or a very limited selection of content for public access
- Key to public access of the whole 1M hours in the BBC archive is finding content:

BBC R&D Multimedia Classification project

- Automatically derives high-level meaning/metadata from audio and video analysis of TV programmes

Comedy

Factual



"I have a cunning plan, sir."



Drama



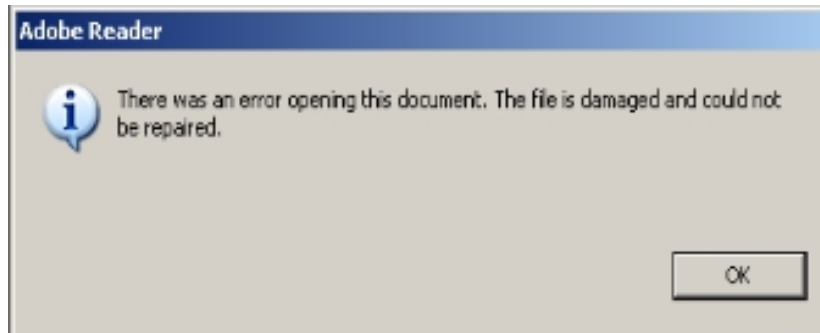
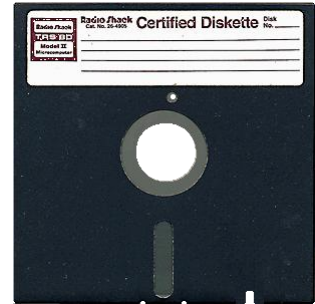
Thriller

- To develop our strategy for managing the BBC's digital archives and enabling public access.

3. Preservation economics in broadcasting

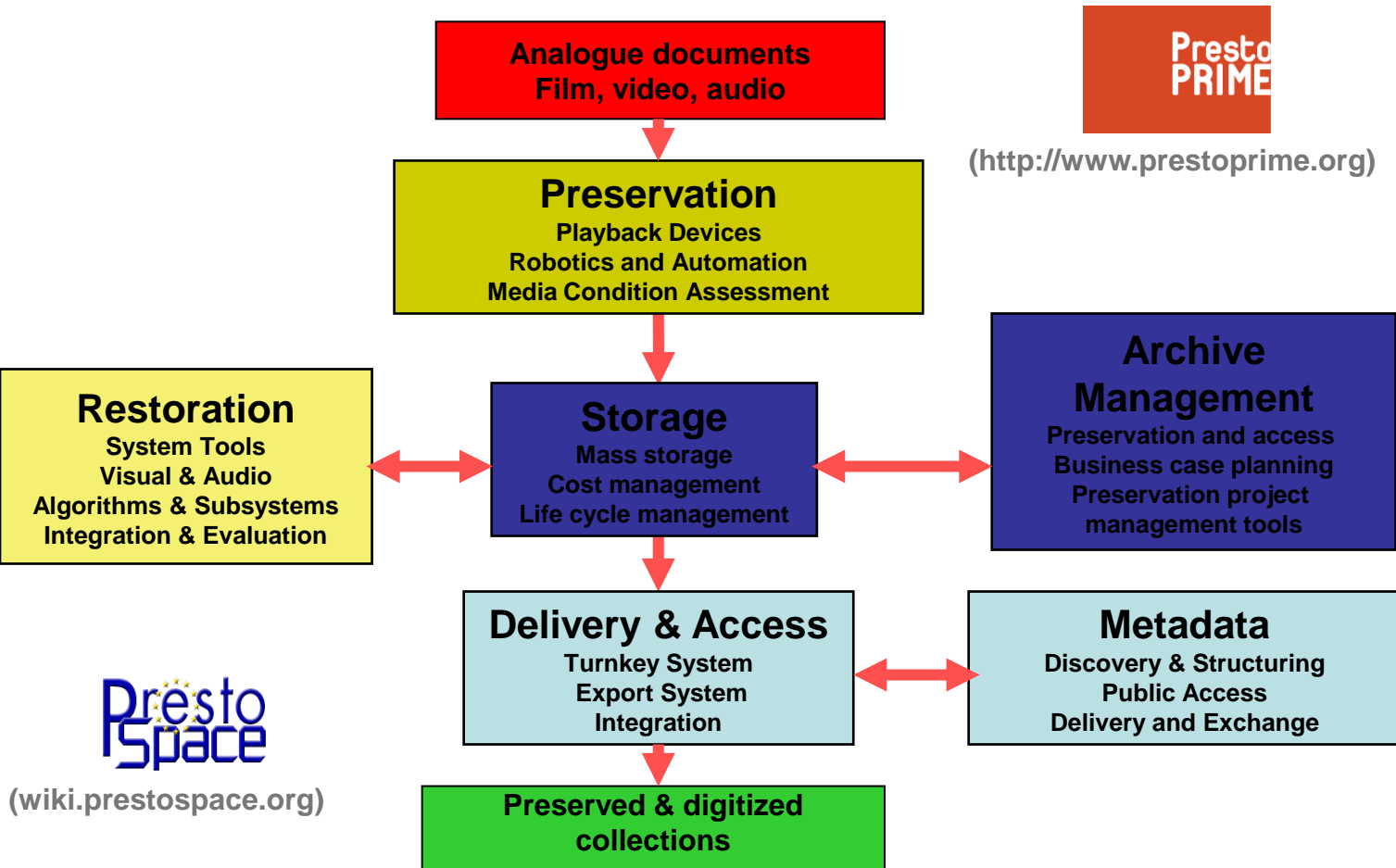
3. Preservation economics in broadcasting

- “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.

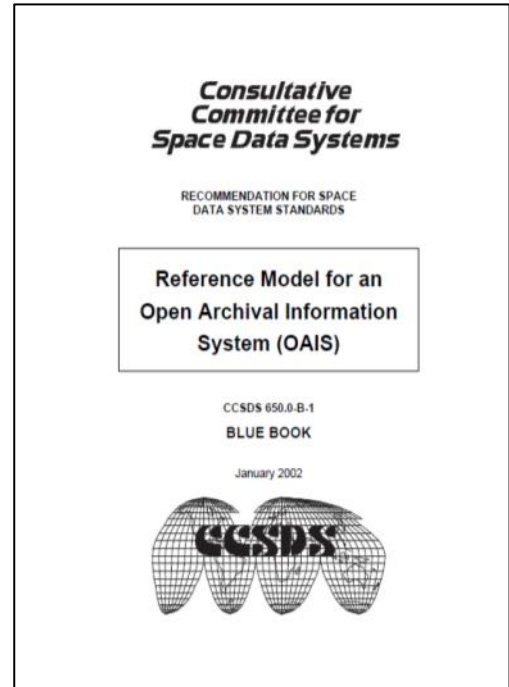
3. Preservation economics in broadcasting: Presto project series



3. Preservation economics in broadcasting

- OAIS Reference Model:
 - A formal process to ensure nothing is forgotten in the operation of a digital archive, such as:
 - Data integrity
 - Security
 - Administration
 - Preservation planning
 - Disaster recovery
 - ...

(<http://public.ccsds.org/publications/archive/650x0b1.pdf>)



3. Preservation economics in broadcasting

Proprietary v Open Standards/Source:

- Proprietary:
 - Supplier support
 - Apparent low cost:
 - But get what you pay for
 - Lock-in to supplier
 - Built-in obsolescence:
 - Future migration problems
 - Especially if supplier out of business.

3. Preservation economics in broadcasting

Proprietary v Open Standards/Source:

- Open Standards/Source:
 - Designed for best archive practice
 - Multiple suppliers
 - Supplier support
 - Can change supplier
 - Can adapt as technology develops
 - Minimal migration issues
 - Can adapt to any special requirements
 - Perceived as amateur.



3. Preservation economics in broadcasting

BBC digital storage requirements:

- 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)
 - 330 PetaBytes (two copies)

3. Preservation economics in broadcasting

Costs:

- SD (new and preservation):
 - 110 TB/wk
 - 130 LTO-4 tapes
 - £4000/week (Aug 09 price)
- HD (when SD production ceases):
 - 200 TB/wk
 - 250 LTO-4 tapes
 - £7000/week (Aug 09 price)
- Total cost of tape alone by 2026:
 - £5.8M (one copy), £11.6M (two copies)

(But data tape prices have been dropping by 50% every two years for a decade)

3. Preservation economics in broadcasting

Video coding to reduce the storage requirements:

- Lossy compression
 - Various options:
 - MPEG1, MPEG2, MPEG4, JPEG, JPEG2000...
 - Typically 10:1 - 40:1 compression
 - Each adds its own type of impairments
 - Coding concatenation builds up impairments:
 - ‘Last Straw’ effect result in sudden appearance of impairments
 - Long-term sustainability of the decoders
 - Will decoders still work on new operating systems?
 - Is emulation a safe way to run decoders?.

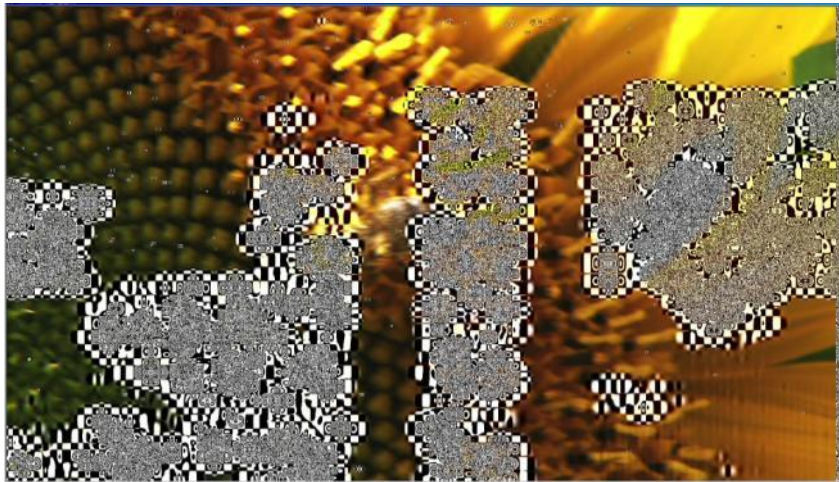
3. Preservation economics in broadcasting

Video coding to reduce the storage requirements:

- Lossless compression:
 - JPEG2000 lossless
 - Typically 2:1 compression ratio
 - Can get original bits back
 - Need to maintain the decoder for the life of the content
 - Ensuring that it works on new operating systems and hardware
 - Perhaps OK if all a/v archives agree to use it.

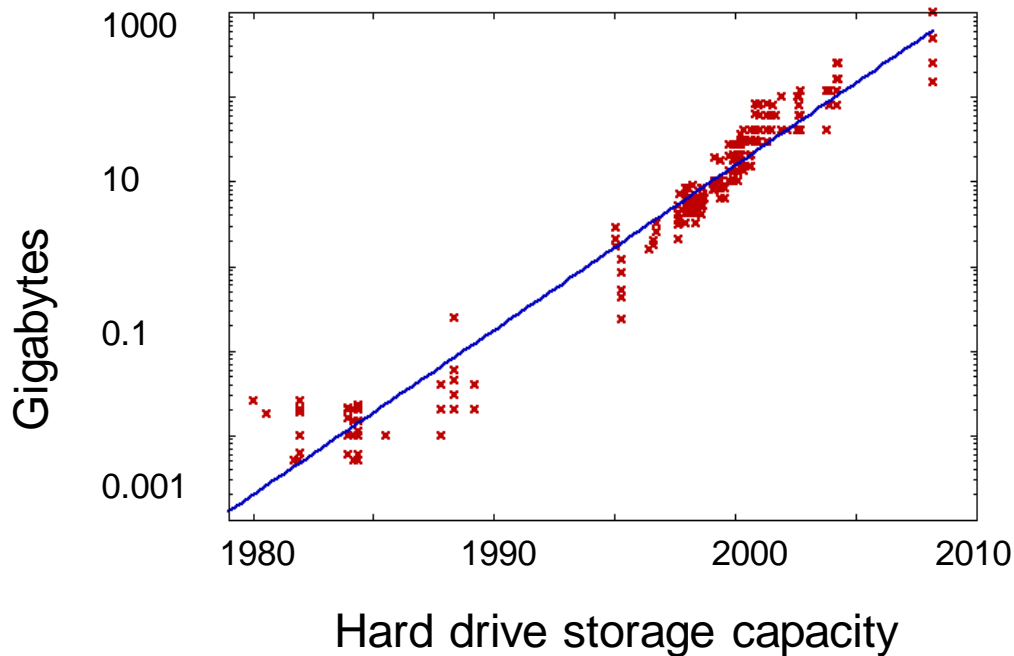
3. Preservation economics in broadcasting

- 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:



3. Preservation economics in broadcasting: Storage costs

- Storage Capacity Goes Up
- Media/Device Cost Goes Down
- Storage is only 10-15% of the total cost of an a/v archive



4. Partnerships

4. Partnerships

- The costs and problems of migrating to digital archives are bringing organisations together to:
 - Understand how to use digital technology and to agree on standards
 - Share information on preservation practices
 - Find new applications for archive content
- An archive should not try to go alone with digital technology!

4. Partnerships: Technology

- Work with the industry:
 - Guide industry to meet archive requirements
- EU Collaborative Framework Projects:
 - e.g. Call 6 for Digital Libraries research projects
- UK TSB-funded projects
- BBC R&D currently working on:
 - PrestoPRIME (EU):
 - Developing recommended practises for “Digital Preservation”
 - Avatar-M (TSB):
 - Developing advanced storage techniques for ExaByte-size files.

4. Partnerships: Preservation standards and practices:

- FIAT/IFTA - International Federation of Television Archives:
 - To provide a forum for exchange of knowledge and experience
 - To promote the development and use of audiovisual archives
 - To establish international standards on key a/v archive issues
- EBU – European Broadcasting Union:
 - P/DATA (Digital Archives Technology Advice)
 - P/SCAIE (Automatic Information Extraction)
- PrestoPRIME Competence Centre:
 - To deliver advanced digital preservation advice
 - Expertise from Europe's major a/v archives
 - Virtual networked facility
 - Expected launch in 2011



4. Partnerships: Applications of archives

BBC is forming an increasing range of partnerships – we are experimenting with how partnerships can work best for each side

- Current application partnerships:
 - British Film Institute:
 - Keep selected TV programmes
 - MOU on archive access
 - Technical knowledge exchange
 - British Library:
 - Keep selected radio programmes
 - MOU on archive access
 - Technical knowledge exchange
 - Open University
 - Do more with archive content
 - Exploring ways to find and tag clips for use in teaching.



The screenshot shows the BBC News website. At the top, there is a search bar and navigation links for 'Text only' and 'Help'. Below the search bar, the 'Press Office' section is highlighted. A 'Press Release' is displayed with the headline 'BBC and BFI announce new partnership to bring archives to life'. The release date is '09.03.2009' and the category is 'BBC: New Media'. The text of the release states: 'The BBC and BFI today announced a partnership to work together to develop plans for increasing public access to their respective audio, film and TV archives.'

Below the press release, there is a 'British Library and BBC sign archive access deal' section. It features a photograph of the British Library building. The text reads: 'The BBC and the British Library are to work together to make their assets more widely available. The arrangement will seek to integrate access to nearly one million hours of BBC TV and radio content and more than 150 million British Library items.'

On the left side of the screenshot, there is a 'BBC NEWS' logo and a navigation menu with categories: 'News Front Page', 'Election 2010', 'World', 'UK', 'England', 'Northern Ireland', 'Scotland', 'Wales', 'Business', 'Health', 'Education', 'Science & Environment', and 'Technology'. At the bottom of the screenshot, there is a 'SEE ALSO' section with links to 'Robots used at £20m library', 'British Library opens music', and '19th Century news goes digital'.

4. Partnerships: Applications of archives

- Partnerships (cont):
 - Arts Council England
 - Help The Arts to use new digital tools:
 - Liberate their archives and Reach audiences with new output
 - Encouraging engagement with the Canvas consortium.
 - SCA/BUFVC* project:
 - Test value of BBC Archive in Higher/Further Education
(* Strategic Content Alliance & British University Film & Video Council)
 - European Solidarity Centre:
 - Potential for external funding and international use of BBC Archive
- Issue for partnerships: Rights clearances.

The Henry Moore
Foundation

BBC Archive and Henry Moore Foundation announce partnership
Treasure trove of art films released online

15.02.2010



Conclusions

- BBC Archives will need at least 330 PetaBytes storage by 2016
- Broadcast industry needs to understand the issues of digital preservation and follow OAIS principles
- Technology choices important for reliable long-term storage
- Archives should work together to present common requirements to archive technology suppliers
- Partnerships can share cost of digital preservation
- Rights clearances for a/v content needs to be simplified to allow innovative re-use.

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

