Presto Centre: tools for audio-visual preservation

Richard Wright
BBC R&D -- PrestoPRIME
DPC – London – 8 April 2011
What I really want to talk about:

• Three kinds of digital preservation
• “preservation and access” = all one concept
  – Requirements for serious access
• PrestoPRIME tools
  – Very brief list
  – One run-through of one online tool
Why Are We Here?

• We have audiovisual content
• The ‘do-nothing option’ fails
• So we take preservation actions
  – Including digitisation
    • Which is expensive
    • And takes us into a new world
• Better, faster, cheaper: Presto, PrestoSpace
• The solution is the new problem: PrestoPRIME
Three Aspects of Digital Preservation

• **Making** analogue content into digital content
  – Digitisation

• **Working** with digital content
  – Digital workflow and processes

• **Preserving** the digital content
  – Digital Preservation
Three Aspects of Digital Preservation

1- Making analogue content into digital content
   - Planning
   - Budget
   - Workflow
   - Standards
   - Rights
   - Result: lots of files

PrestoSpace information online:
//wiki.prestospace.org/  //digitalpreservation.ssl.co.uk/

Now: revised for PrestoCentre
Preservation Guide - Introduction

March 2011: Screening the Future 2011 ---- Major conference launching the PrestoCentre

New Strategies and Challenges in Audiovisual Archiving

Monday, March 14, 2011 at 9:30 AM - Tuesday, March 15, 2011 at 4:30 PM (GMT+0100)
Hilversum, Netherland ---- Full details here: http://prestocentre.eu/

January 2009: New Project: PrestoPRIME It's not Presto3 or Son of Son of Presto but it is related: life after digitisation. After years of digitising audiovisual content, what do we do with all those files to keep them safe, and working properly? All the answers from PrestoPRIME

March 2008: PrestoSpace Integrated Website: All the online results from PrestoSpace in one place, fully integrated! http://digitalpreservation.ssl.co.uk/
The Rolling News, a pushdown list of news flashes

General Guide to Audiovisual Preservation

Is this your problem?

If you have audiovisual media, it needs maintenance – or you will lose it. This guide shows how to:
- conserve old formats
- digitise for transfer to new formats
- create digital file formats
A Web site for archive owners, academics and the public alike, which provides a first point-of-call for information on the migration and digital preservation of audio/visual works. The site is a repository of the information and experience gained in the PrestoSpace project. You will find information ranging from the tutorial to the academic on planning and executing a digital preservation project and the technologies and processes involved. You will also find training videos, management tools and links to online resources and events in the field of digital preservation.

Highlights

- **The SAM analysis tools**: Tools to help you plan and manage a digital preservation exercise.
- **Database of audiovisual media**: Common audio and video formats found in European archives.
- **Training Videos**: Videos showing how to use early videotape equipment.

Events

**Final PrestoSpace Workshop, Rome**

See ALL the PrestoSpace technology, from new machines for audio, video, and film digitisation, to the Turnkey System for hosting your own Digital Audiovisual Archive. More than 10 different demonstrations will run simultaneously presenting the PrestoSpace technology including audio, film and video scanning tools, to tools, guidelines, and services to manage the migration process and the storage, to audio, video, and film restoration tools, and eventually the Publication Platform and the Turnkey System to give access to audiovisual contents.
Three Aspects of Digital Preservation

2- Working with digital content (lots of files)
   - Management
   - DAM/MAM
   - Repository
   - Storage
   - Metadata
   - digital library technology
   - Access
   - Rights
Three Aspects of Digital Preservation

• **3- Preserving** the digital content
• Keeping the data ‘forever’
• Coping with obsolescence
• Migration
• Emulation
• Standards: OAIS and all that
• Planning and strategy
Conservation of Analogue Content

- Handling, packaging and storing
- Environmental conditions
- Protecting the masters
- Condition monitoring
Preservation of Digital Content in Annual Preservation Status Report

• Handling, packaging and storing
• Handling: fixity check
• Packaging:
  – Wrapper formats
  – Encodings
  – Embedded metadata
• Storing:
  – Storage technology
  – Cost models: century store; “forever” costs
• It’s all on the PrestoCentre website
Environmental conditions

- The *social, political and economic* environment of a Trusted Digital Repository
- A lot of work has been done: TRAC, DRAMBORA
- **TRAC Criteria Documents**
  - A1.2 Contingency plans, succession plans, escrow arrangements (as appropriate)
  - A3.1 Definition of designated community(ies), and policy relating to service levels
  - A3.3 Policies relating to legal permissions
  - A3.5 Policies and procedures relating to feedback
  - A4.3 Financial procedures
  - A5.5 Policies/procedures relating to challenges to rights
More TRAC

B1  Procedures related to ingest
B2.10  Process for testing understandability
B4.1  Preservation strategies
B4.2  Storage/migration strategies
B6.2  Policy for recording access actions
B6.4  Policy for access
C1.7  Processes for media change
C1.8  Change management process
C1.9  Critical change test process
C1.10  Security update process
C2.1  Process to monitor required changes to hardware
C2.2  Process to monitor required changes to software
C3.4  Disaster plans
Protecting the masters

• Analogue content needed viewing proxies
  – For protecting of the masters
• Digital content needs viewing proxies
  – For protecting the bandwidth!
• Three-level approach:
  – Master
  – Mezzanine: the most efficient coding for generating new proxies
  – Viewing proxies (could be multiple quality levels)
Condition monitoring

• Analogue: check the stock on the shelves
• Digital: check the stock (test the files)
  – Fixity check
  – Beyond fixity:
    • Identification: fingerprint technology
    • Quality analysis: technical measurements on the audio and video signals contained in the files
  – AND – check the ‘shelves’: test the storage system
  – Monitoring needs cost modelling, to set a cost-effective strategy
Lots to manage!

- Lots of help from the PrestoCentre
- Including risk and cost analysis and strategy and simulation tools from PrestoPRIME
  – and I will demonstrate one

but first: requirements for access
Four requirements for sensible access

- Granularity
- Navigation
- Reference and Citation
- Annotation
Granularity - division into meaningful units

- Keyframes
- Other methods to represent video

- and audio:
Navigation

• "Click and play" on visual representation of the meaningful units
Reference and Citation

• the core requirement for scholarly discourse
  – along with a major change in attitude!
• Needs a permanent place for “things to be”
  – Hence the need for stable audiovisual collections

“Hamlet, for example, is comparable to Saxo Grammaticus' Gesta Danorum. [citation needed]
King Lear is based on King Leir in Historia Regum Britanniae by Geoffrey of Monmouth, retold in 1587 by Raphael Holinshed. [citation needed]"
Annotation

- the core requirement for social web = interactivity
- individual interacts with content
- individuals interact with other individuals
Back to preservation tools: PrestoPRIME Technology

• Digital Preservation Demonstration systems
  – Our own, and Ex Libris Rosetta
  – Supporting broadcast files and metadata: MXF

• Lots of work on metadata
  – Collecting, ‘gardening’, mapping
  – Document on audiovisual preservation metadata
  – W3C Media Annotation and Fragments WGs
  – Rights ontology -- for rights automation
PrestoCentre

• European Networked Centre of Competence
  – Prestocentre.eu

• The plan: membership and other ways to be self-supporting
  – So it doesn’t end when the PrestoPRIME project ends
RT @NeelieKroesEU: #EU countries need to comply w/ EU rules on #AV #media services. Wrote to 16 MS today to make sure they do. bi ...
20 hours 47 min ago
Follow us on Twitter

News

Screening the Future video registration now online
MARCH 18
On March 14th and 15th, the official launch of PrestoCentre took place at the Screening the Future Conference at the Netherlands Institute for Sound and Vision, in Hilversum, the Netherlands.
Over 200...

PrestoCentre Blog

The challenge of finding video on Google
MARCH 10
Claire Harvey - The past few years have seen the AV archive industry mature and progress at a significant pace. Since the advent of the Presto projects at the beginning of the century, the industry has moved from being manual and analogue...
The challenge of finding video on Google

MARCH 10

Claire Harvey - The past few years have seen the AV archive industry mature and progress at a significant pace. Since the advent of the Presto projects at the beginning of the century, the industry has moved from being manual and analogue...
And now:

one PrestoPRIME tool

• A model for storage systems, to calculate
  – Cost
  – Risk
  – Loss
  – And compare what-if scenarios
Login

Please enter your username in the box below.

RichardBBC

Log In

If you do not already have a username then you need to create one so that any model data you create can be saved.

This can be done simply by choosing a name of anything you like that is unique to you, putting it in the box above, then pressing ‘Log In’. No password is needed.

Please note that other people will be able to see your saved data if they know or guess your username.
Storage Systems

Found 4 storage systems. Add...

HDD in servers  read-only
Migration required every 4 years.
Running Costs
Access: €0.1 per GB
Storage: €1 per GB per year
Corruption Rates
Access: avg. 1 in 500 files
Latent: avg. 1 in 750 files per year

HDD on shelves  read-only
Migration required every 4 years.
Running Costs
Access: €1 per GB
Storage: €0.25 per GB per year
Corruption Rates
Access: avg. 1 in 100 files
Latent: avg. 1 in 500 files per year

Data tapes in a robot  read-only
Migration required every 6 years.
Running Costs
Access: €0.2 per GB
Storage: €0.4 per GB per year
Corruption Rates
Access: avg. 1 in 10^4 files
Latent: avg. 1 in 10^5 files per year

Data tapes on shelves  read-only
Migration required every 6 years.
Running Costs
Access: €1 per GB
Storage Systems

HDD in servers
Migration required every 4 years. **Running Costs**
Access: €0.1 per GB
Storage: €1 per GB per year

**Corruption Rates**
Access: avg. 1 in 500 files
Latent: avg. 1 in 750 files per year

HDD on shelves
Migration required every 4 years. **Running Costs**
Access: €1 per GB
Storage: €0.25 per GB per year

**Corruption Rates**
Access: avg. 1 in 100 files
Latent: avg. 1 in 500 files per year
More Storage Systems

Data tapes in a robot
Migration required every 6 years. **Running Costs**
Access: €0.2 per GB
Storage: €0.4 per GB per year

**Corruption Rates**
Access: avg. 1 in 1x10^4 files
Latent: avg. 1 in 1x10^5 files per year

Data tapes on shelves
Migration required every 6 years. **Running Costs**
Access: €1 per GB
Storage: €0.1 per GB per year

**Corruption Rates**
Access: avg. 1 in 1x10^4 files
Latent: avg. 1 in 1x10^5 files per year
Storage Configurations

Found 3 storage configurations. Add...

**Disk with Tape**  read-only
- System 1: HDD in servers
  - Files accessed avg of 0.25 times per year, staying constant.
  - Scrubbing every 1 year(s).
- System 2: Data tapes in a robot
  - Files accessed avg of 0 times per year, staying constant.
  - Scrubbing every 3 year(s).

**Disks on Shelves**  read-only
- System 1: HDD on shelves
  - Files accessed avg of 0.25 times per year, staying constant.
  - Scrubbing every 5 year(s).
- System 2: HDD on shelves
  - Files accessed avg of 0 times per year, staying constant.
  - Scrubbing every 5 year(s).

**tape on shelves**
- System 1: Data tapes on shelves
  - Files accessed avg of 2 times per year, staying constant.
  - Scrubbing every 100 year(s).
- System 2: Data tapes on shelves
  - Files accessed avg of 0.01 times per year, staying constant.
  - Scrubbing every 100 year(s).
Found 3 storage configurations. Add...

Disk with Tape
System 1: HDD in servers
Files accessed avg of 0.25 times per year, staying constant
Scrubbing every 1 year(s)
System 2: Data tapes in a robot
Files accessed avg of 0 times per year, staying constant
Scrubbing every 3 year(s)
File Collections

Found 1 file collection. Add...

Default File Collection  read-only
Length of cost/loss projection is 25 year(s).
Files
100 thousand initially, staying constant.
Average File Size
25 GB.
File Collections

- Found 1 file collection. Add...
- read-only
- Default File Collection
- Length of cost/loss projection is 25 year(s).
- Files
- 100 thousand initially, staying constant.
- Average File Size
- 25 GB.
Plans

Found 3 plans. Add...

Default Plan
File Collection: Default File Collection
25 year lifetime. 100 files, avg. 25 GB in size.
Storage Configuration: Disks on Shelves
Uses HDD on shelves and HDD on shelves systems.

Evaluate

Tape on shelves
File Collection: Default File Collection
25 year lifetime. 100 files, avg. 25 GB in size.
Storage Configuration: tape on shelves
Uses Data tapes on shelves and Data tapes on shelves systems.

Edit Delete Evaluate

Disk and Tape
File Collection: Default File Collection
25 year lifetime. 100 files, avg. 25 GB in size.
Storage Configuration: Disk with Tape
Uses HDD in servers and Data tapes in a robot systems.

Edit Delete Evaluate
Plans

Found 3 plans. Add...

Disk and Tape edit Delete Evaluate

File Collection: Default File Collection
25 year lifetime. 100 files, avg. 25 GB in size.

Storage Configuration: Disk with Tape
Uses HDD in servers and Data tapes in a robot systems.
Thank You

• Storage model: http://prestoprime.it-innovation.soton.ac.uk/planning-tool/
• PrestoPRIME prestoprime.eu
• PrestoCentre prestocentre.eu
• BBC R&D bbc.co.uk/rd
• Richard Wright richard.wright@bbc.co.uk