PRESTO – Preservation Technologies for European Broadcast Archives

IST-1999-20013

D7.1 Metadata Reference Process Analysis

DOCUMENT IDENTIFIER	PRESTO-W7.1-BBC-020509 Metadata Reference Process
DATE	09 May 2002

ABSTRACT: A detailed definition of the workflow for optimised archive preservation. The automated management of metadata requires a process control system (workflow management software), which provides an efficient central controller for an entire preservation process.

AUTHOR, COMPANY	R Wright BBC
INTERNAL REVIEWER	
WORKPACKAGE / TASK	WP7 / T1

DOCUMENT HISTORY

Release	Date	Reason of change	Status	Distribution
1.0	17.04.2002	First draft	Draft	Confidential
1.1	06.05.2002	Revision	Complete	Public
1.2	09.05.2002	Added 3.5 = practical examples	Complete	Public

Table of Contents

1. Summary

The workflow for the “preservation transfer” of archive materials, as documented in D3.2, is analysed to develop a **reference transfer process** that makes the overall process as efficient as possible. The analysis looks at the whole picture, maps the workflow, documents the allocation of time and effort, and develops an optimised model for archive preservation, incorporating key-links also developed by PRESTO, and incorporating new technology outside PRESTO that may not be used (or widely used) in preservation work.

2. Where does preservation work begin and end?

As discussed in the following section, preservation has many aspects. The purpose of this chapter is to establish the significance of 'transfers' in the whole preservation picture. Section 2.1 describes all the main aspects of preservation. Section 2.2 distinguishes between preservation transfers and other archive transfers, and finally Section 2.3 distinguishes between preservation and restoration.

2.1. Media lifecycle

2.1.1. Content vs carrier

Unlike museums or conventional archives, broadcast archives have generally accepted the distinction between the audiovisual content (for example: radio or television programme, musical performance), and the physical carrier of than content (for example: videotape, reel of film, audio cassette or CD). Further, they have accepted that the carrier is merely a means to an end, and not an artefact to be preserved in its own right.

This frees a broadcast archive from the need to preserve physical objects, and allows it to concentrate on the most efficient method of preserving the content. Efficiency is measured on two main dimensions: cost and usability. In this case "usability" denotes the ability of the broadcasting organisation to use the carrier held in the archive. One major reason for 'preservation projects' is not that carriers are deteriorating, but that they are obsolete technology. In practice, all media carriers are deteriorating – the only difference is rate of decay. However format obsolescence is as much a driver of preservation work as is the issue of deterioration – as documented in Deliverable 3.1.

Whether the determining factor is decay or obsolescence, the result is the same: all carriers have a finite usable lifetime. The essence of preservation work is the transfer of content from one carrier to another. Hence there is a tendency to think of these transfers as constituting preservation, whereas properly they are not the whole preservation picture at all, but only a vital stage.

Amongst the various possible preservation transfers, one is distinctive: the transfer from an analogue to a digital carrier. This does not break the cycle of recurrent needs for transfer, but does alter the whole technological and economic character of subsequent transfers. Analogue-to-analogue or analogue-to-digital transfer is characteristically expensive, manually-intensive, involves potential loss or alteration of the original signal (content) and requires the monitoring, intervention and judgement of skilled staff. Digital-to-digital transfers can largely be automated, with robotic handling, completely automated transfers, and perfect replication of the original content.

The main focus of PRESTO is on the analogue-to-digital preservation transfer, and ways to make it faster and cheaper while maintaining broadcast standard content quality.

2.1.2. Analogue lifecycle

The lifecycle of an analogue carrier is from the time it is created (ie has content written on it) to the time its content is transferred to another carrier. The lifecycle costs associated with the carrier have the following main components:

- Creation: media cost; cost of equipment and operator needed to write the content onto the carrier

- Accession: costs associated with identifying and cataloguing the carrier to bring it into the archive
- Storage: shelf space costs
- Maintenance: environmental conditioning and any other costs required to maintain the carrier in good condition; production of viewing copies and new 'secondary masters' (to replace worn-out, lost or damaged copies) are maintenance costs
- Access (retrieval): cost of method used to get the item from the archive into the hands of an archive user, including "format transfer" or production of a "user copy" if required
- Usage (circulation): costs associated with tracking use of the item, and retrieving it (returning it to the archive) if necessary
- Clearance: costs associated with allowing the content to be used for a business purpose
- Transfer to new carrier: the final transfer to a new carrier

The total lifecycle costs will then vary from item to item, because maintenance, access, usage and clearance costs will depend upon the how often a particular item is circulated.

2.1.3. Digital lifecycle

There are two forms of digital carrier:

1. Those which are very much the same as an analogue carrier, except using a digital method of writing the content onto the carrier. Examples are minidisc and CD in audio, and digital videotape (D3, D5, Digibeta) in video.
2. Those which use file formats on servers (hard disc drives) and/or datatape, with electronic distribution of the content over networks.

The lifecycle for a digital carrier of Type 1 is very much the same as for analogue carriers, except that transfers for maintenance, access and final transfer to a new carrier will be cheaper than for analogue – as no operator intervention (beyond physical handling) is needed to make a perfect 'clone' of the digital original.

The lifecycle for Type 2 is simpler than for analogue, because if delivery is by a network then there is no physical item to be tracked and returned to the archive. However the overall equipment and especially the storage and delivery infrastructure may be very expensive in comparison with Type 1 working.

2.1.4. Preservation transfers: the key "preservation process"

Preservation has many aspects, and in its widest definition encompasses all processes necessary to the sustained existence of the content over an indefinite period – a period which for audiovisual material will involve holding the content on a succession of carriers.

But as in a relay race, a critical moment is when the baton is passed – literally from carrier to carrier. The process of transferring content from an old to a new carrier is of special significance to broadcast archives for several reasons:

- **Time and effort:** the transfer of thousands of items – in some cases hundreds of thousands – can take many years. The transfer is a major activity, and is literally a race against time as the original carriers decay or fade into obsolescence.
- **Possibility of loss:** reading and digitising of analogue data can be done well or poorly. There are many ways to do a less than optimum transfer, and once transferred any loss of original quality may be a permanent loss. Even if original

carriers are kept, their subsequent deterioration combined with lack of funds for re-transfers means that the first mass transfer off analogue is likely to be the only opportunity to capture the original with highest achievable fidelity.

- **Cost:** broadcast archives are facing the need to take action on up to 75% of their current carriers. Costs for major transfers are not built into standard operating budgets, so special measures are needed to get funding for the transfers.
- **Technology change:** broadcast technology is rapidly changing, and updating the archive needs to transfer from obsolete carriers to those which will be most useful to current and foreseen production. This task is not at all simple when production areas themselves are struggling with technology choices and strategies.
- **New opportunities:** new technology opens doors to new service and commerce opportunities. The major investment required for preservation transfers is for most archives the largest monetary investment these archives will ever receive. It is the goal of all archivists to use this money to best possible advantage in terms of making farsighted technology choices.

2.2. Standard archive transfers

Archives make transfers all the time. This section attempts to distinguish such 'standard' activity from the special projects needed for preservation transfers. The basic difference is between mass transfers and one-off requirements. The significant savings that can be achieved through transfer automation also require significant capital investment, staffing, and volume of material. The following is a list of the main 'standard' operations that are similar to preservation transfers, but differ in being one-off, on-demand, low-volume operations.

2.2.1. On-demand

Multimedia archives hold material on various formats, and when an item is required which is on the 'wrong' format, it is usual, especially for audio and video but not for film, to make an 'on the spot' transfer to the required format. This is daily activity in broadcast archives.

The main problems with using an on-demand approach to preservation transfers are:

- **Time:** the old format may be decaying too rapidly, or becoming obsolete too rapidly, to allow an on-demand process to run its course.
- **Cost:** although transfer on-demand may appear to be free, it is inherently the least automated and therefore most expensive way to do preservation transfers, if full costs are known and appreciated.

2.2.2. Telecine

Television does not broadcast directly from film, though 20 or more years ago it was common practice to use 'live telecine' at time of broadcast. Now, anything held on film must at some stage be converted to video before it can be used. It is commonplace for archives to do this work, either internally or outsourced.

Telecine transfers are not preservation work, because the original film is almost always kept. One major savings of adopting a 'digital film' approach would be the elimination, forever, of these telecine operations.

2.2.3. Digitisation

It is common for broadcasters to operate a mixture of analogue videotape, digital videotape and server-based methods in production and archiving. In such situations, media is often

written to and read from digital processing stations, such as non-linear videotape editors or audio workstations. These digitisations are a form of transfer, and like telecine are a necessary, expensive, time-consuming and consequently frustrating means to an end. In a fully networked, server-based environment these transfers would be eliminated.

Digitisation transfers are not preservation work for the same reasons as for telecine: the original is kept, the time and effort put into the transfer is ultimately completely lost in terms of not achieving any permanent result – only enabling a specific process (such as an edit) to take place.

2.2.4. Compression

Modern broadcasting uses a variety of versions of 'content', from highest available quality down through various lower datarate encodings. The lower datarate versions (proxies) are used for browsing, for digital broadcasting, and for websites. The process of making such proxies is very similar to digitisation in terms of time, effort and equipment. A difference is that there will be outputs (proxies) that can be permanently kept. These proxies will not replace the originals because they are of lower quality, and so these transfers are not preservation operations.

2.3. Restoration

A basic principle of multimedia preservation is 'warts and all': preserving the original signal (content) as nearly as possible in the exact state as represented on the carrier at the time of a preservation transfer. This approach is in line with a general archive principle of 'preserving the original artefact'.

For multimedia, the true 'original artefact' never enters the archive. The real original is the performance or production that was taped, filmed or recorded to make the original carrier. We can never capture that original, but we do know the difference between how the world looks and sounds, and a faded, scratched, noisy or otherwise degraded signal which is now all we have to represent the original event.

Signal processing can alleviate many forms of signal impairment. Such work, if successful, can make content closer (in perceptual terms) to the original event. It can certainly make signals more acceptable for broadcasting, by eliminating or reducing scratches, clicks, obvious masking noises and other problems.

Signal processing can also remove content rather than noise, and otherwise alter or even degrade rather than 'restoring'. A major problem with restoration signal processing is that is always takes human judgement to guide the processing, making restoration work time-consuming and expensive.

For the purposes of this document, 'preservation transfer' covers only the processing needed to capture the signal currently on the carrier, not any of the further processing steps used to improve upon that signal. All such attempts to clean-up, correct or cosmetically enhance the signal will be considered restoration and are outside the scope of this document.

A tricky case arises in video. It is common for a single line of video to be lost in playback of videotape. Of course multiple lines and whole frames can also be lost. Most modern videotape players have 'concealment', so that if a line is misread, the preceeding line is played out twice, to not make a gap in the picture. Because this is a standard, automatic and accepted aspect of videotape replay, archive preservation involving videotape will include simple concealment – even though strictly speaking concealment is a primitive form of signal restoration.

3. Basic workflow for Preservation Transfers

3.1. PRESTO generic preservation process

The preservation process as considered in the PRESTO project is given by a nominal chain consisting of four basic stages:

- Composition
- Digitisation
- New media creation
- Archive update

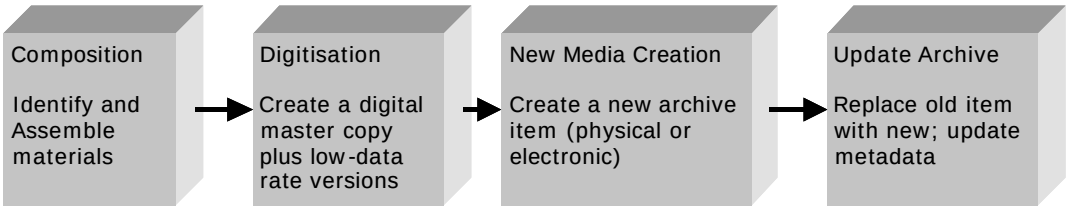

Figure 1: Reference process

At each step, a number of actions has to be performed both at media processing and at metadata processing level. They are summarized by the following figures.

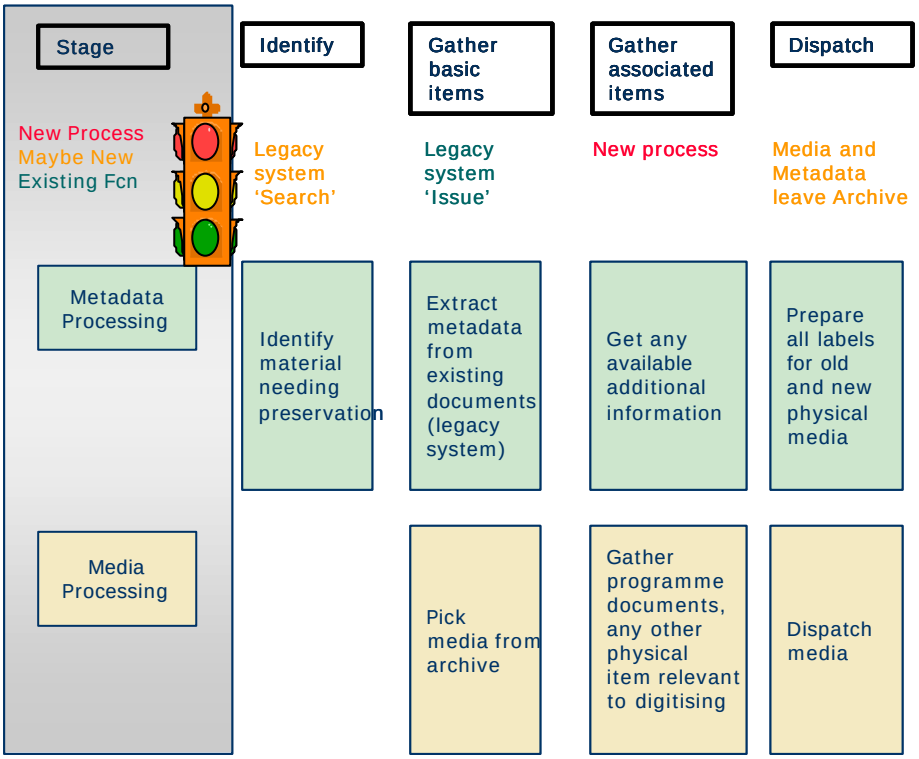

Figure 2: Composition

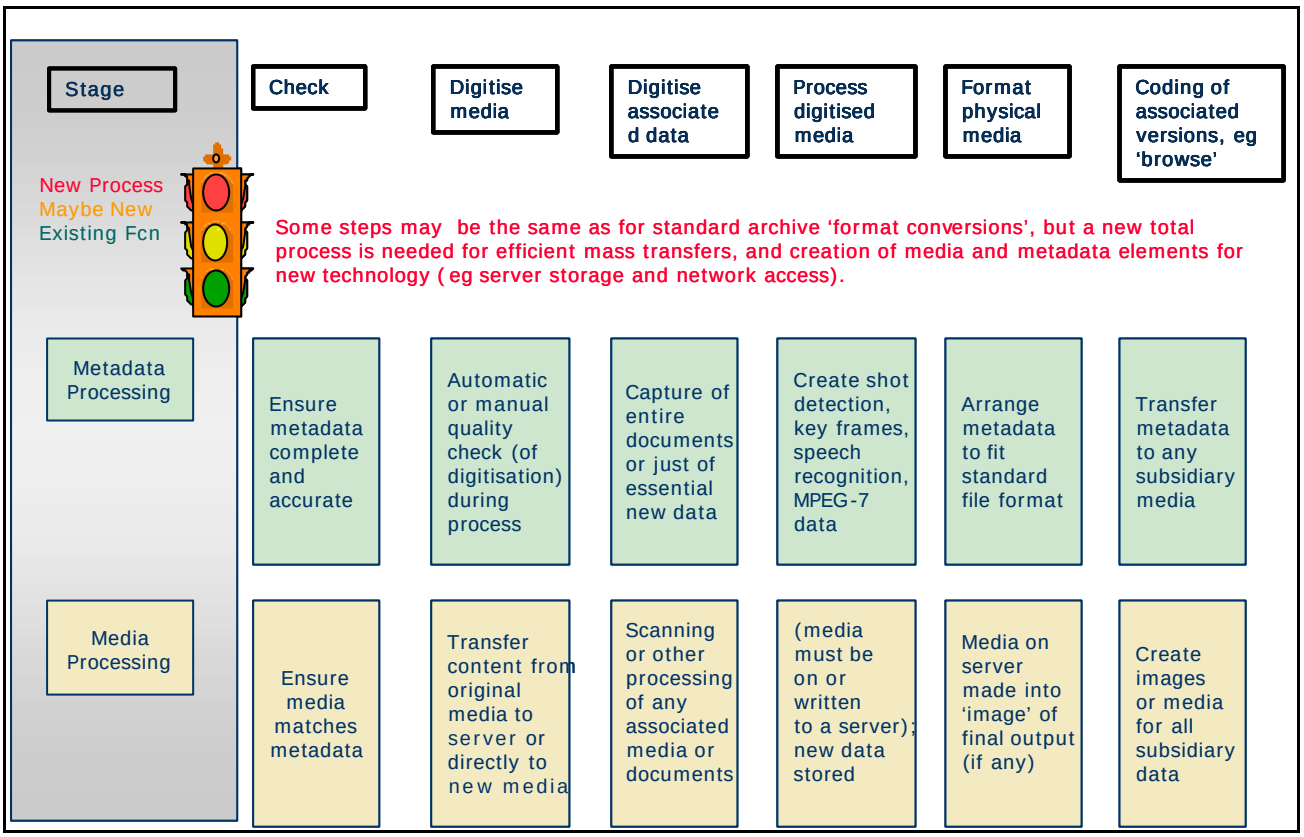

Figure 3: Digitisation

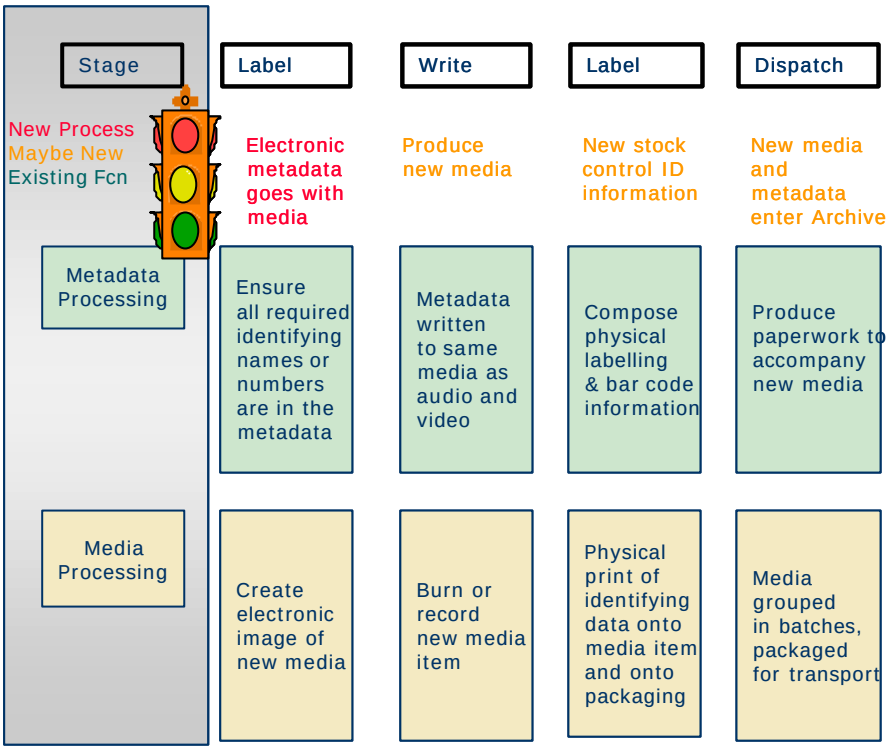

Figure 4: New media creation

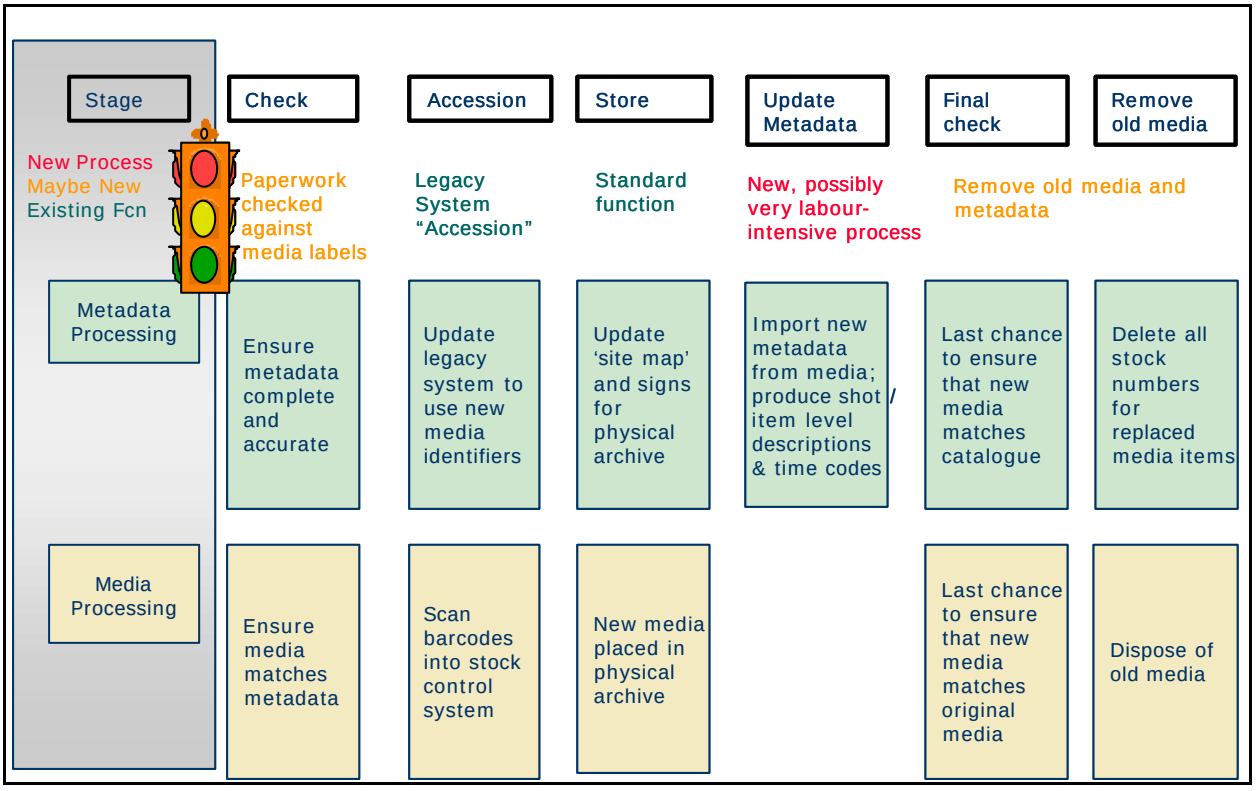

Figure 5: Update of physical archive

3.2. Combined Workflow

The stages listed in section 3.1 above can be simplified and combined into one workflow. The following diagram is a bit complex at first reading, but it is has the useful feature of showing all the major steps in a preservation transfer process on one diagram.

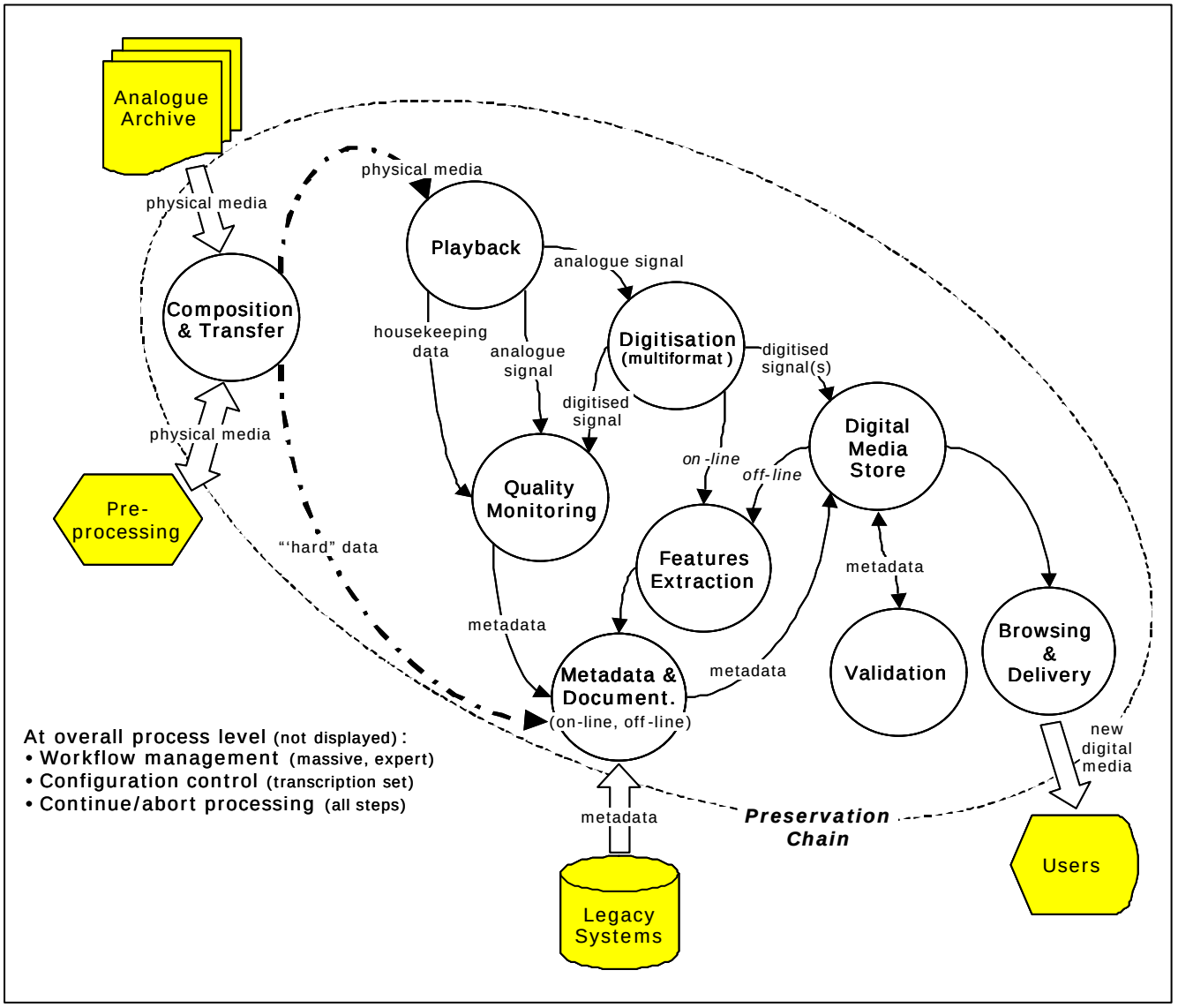

Figure 6: Preservation Transfer Workflow

How to use Figure 6: Preservation Transfer Workflow **to save time and money and improve quality**

- Adapt the diagram to your own process workflow
- Look at every arrow: what metadata is associated with every arrow?
- Where does the metadata come from, and where does it go?
- How can it be automated?

The optimum process: metadata comes into the system in electronic form, media are barcoded when first handled, a computer programme produces ALL further metadata and the entire workflow is tracked by the programme to ensure nothing is overlooked.

3.3. Time and effort

In addition to fully automated metadata handling and process tracking, the following points are important to effective allocation of time and effort:

- Handling physical media
  - **Reducing number of times an item is handled**
  - **Reducing/consolidating operations preformed each time an item is handled**

    Every time an item is 'picked up and put down', there is cost. If two versions of an item are needed (ie full quality and browse, or master copy and lending copy), an efficient process makes them both at the same time, not sequentially.
  - **Number of transcription stations, their cost and usage**

    The number of workstations clearly relates to the time a project will take, but there are other considerations. An investment in automated workflow will require various 'central resources': software to manage the metadata and track the workflow, possibly a network and server, supervisory and support staff (in addition to the transfer operation staff). Any investment in a 'central resource' has a cost that is an overhead on the overall project. The more workstations, the lower the overhead per individual workstation.

    Capital investment is 'written off' over a number of years, anything from three to ten being typical. Workstations that are used 24 hours per day, every day of the year earn a greater return on capital investment than for equipment which is less fully utilised.

- Monitoring
  - **Number of signals or media items handled per operator**

    Current state-of-the-art is for one operator to perform five simultaneous transfers (RAI Radio project). In on-demand audio transfers, and generally in video transfers, an operator does one transfer at a time.

    There is a very substantial quality issue associated with having an operator attempt to monitor multiple transfers. The answer to maintaining quality despite not having the full-time attention of the human operator is through the use of sophisticated monitoring equipment and software. PRESTO has extended the state-of-the-art in this area, with software libraries for automation of both audio and video quality monitoring, and with dedicated hardware for equipment monitoring.

    The use of automation for quality monitoring is another area where capital investment is needed in order to obtain greater workflow efficiency.

3.4. Costs

The costs of preservation are really lifecycle costs, covering every aspect of creation, storage, use and ultimately re-copying to new media. As discussed, a key stage is the transfer from old to new media, especially the unique occasion of transfer from analogue to digital.

3.4.1. Standard Costs

The costs of such preservation transfers are still by no means easy to establish. It is common to simply quote the cost of the operator and the new media, but there are many other sources of cost in a full 'preservation transfer' project. The PRESTO project has examines these costs. The following list covers the full range of costs for most projects.

Typically, the transfer is less than half the total cost. This is a very rough statement because it is a rare project where total costs are understood. Also typically, transfer costs are in the area of 100 to 200 Euros per hour for audio and video, and roughly 1000 Euros per hour for film. As stated, total 'preservation transfer project' costs are at least twice as high.

Standard cost factors in preservation transfers are:

- Technical work – the actual transfer
- quality control
- metadata
- media
- storage
- media handling
- capital
- maintenance
- transport

3.4.2. Opportunity Links

The irritating thing about costs is that the answers to most questions about costs begin with the words "It depends". A primary dependency regards whether a step is taken in isolation, or combined with other work already deemed necessary. As an example, if a house already has scaffolding to fix the roof, that becomes the cheapest time to also repair the chimney.

Preservation transfers create a window of opportunity to reduce the cost of other processes. This statement is especially true for transfer workflow which is based on digitisation to a server, and electronically-managed processing (as discussed in 4.2, below).

Once material is on a server, there are a range of actions which can be taken to prepare media for new services (eg online browsing) and new business opportunities (eg e-commerce).

- Computer processing can be applied to extract metadata to improve access to the material. An example is speech recognition to produce text data which may have too many errors to be usable as a transcript, but have sufficient information to aid retrieval via text search. A range of technical metadata parameters for audio and video processing are possible, and their description is being standardised as MPEG-7 [1].
- Multiple versions of the material can be produced in various qualities: full, edit, broadcast, browse, internet
- Video can be automatically divided into shots, with one of more 'thumbnail' images extracted for each shot. These thumbnails can be added to the catalogue data for the material, improving media description, reducing the manual effort needed to type-in the basic visual description of video material, and supporting access to the material at the

'item' rather than 'full programme' level. Item-level access is essential for electronic access, as download of full programmes is a very inefficient way to get to individual items.

- Audio can also be divided – into speech, music, applause, other – and even divided speaker-by-speaker. Again, icons can be associated with individual sections, to support item-level access.

3.5. Example metadata workflows

3.5.1. BBC: Audio preservation

The BBC Radio 1 Archive, which started in the 1960's, comprises over 40,000 tapes and DATs (14,500 hours). The majority of the tapes are unique master recordings of pop music and interviews. The catalogue data was on legacy database systems, and incomplete. Given the historic and commercial value of the archive it was decided to provide improved access and protection by digitisation. The main transfer, to CD-R & DVD-R, started in January 2000.

The Transfer Process

Overview

All the work is being undertaken in-house at BBC Maida Vale, which is an apt location, since many of the recordings were originally produced in the music studios.

The specially adapted catalogue and transfer areas at Maida Vale, which employs a total of 19 people, has been divided into three areas: -

- A cataloguing area where tapes arrive and are checked against the databases.
- A transfer area for the tape/DAT transfer, with six networked SADiE 24/96s to produce CD-R copies.
- A second transfer area with a further two SADiEs where DAT tapes are transferred along with three workstations and RAIDs to temporarily hold the BWAV files delivered over the network. A DVD-R burner is also attached to each workstation to make the DVD-R compilations of the BWAV files.

The total RAID storage of 208 GB is filled every three to four days, and up to 120 CD-Rs are produced daily.

Catalogue &data functions

The whole data collection and transfer process is driven by specially written software called RAPPER (Radio and Programme Preservation, Editing and Retrieval). The main points are that data is only entered once, the tapes CD-R's and DVD-R's are all tracked by bar code, and the metadata is automatically added to the BWAV header.

The process and data flow is shown below in Figure 7:

1) The tapes, which had never been catalogued and for which there was no other source of data were manually indexed and information entered into an Access database.
2) For most of the material which had been broadcast there was a log of the basic information in a legacy database called Romeo. A dump of this was obtained, filtered and transferred to an intermediate database called Juliette, where corrections and additions were made.
3) All of this was used to form the core data in the RAPPER system.

4) This information was used to create requisition lists
5) These were Emailed to the Archive at BBC Yalding house.
6) Accession numbers in the form of bar codes are added to each tape.
7) tapes are then despatched to Maida Vale, where they are checked against their bar code in the catalogue area and any corrections are made to the RAPPER database.
8) Batches of tapes, with worksheets are passed to the transfer area, where four tapes are simultaneously loaded into each SADiE workstation and the audio CD-R is burned. Constant audio monitoring is carried out on a rotational basis.
9) The BWAV file is sent over a local network to the RAID array and information from the RAPPER database is used automatically to populate the BWAV header (including USID), before being burned onto DVD-R.
10) The additional data, which cannot fit into the BWAV header, is also kept in a readme file at the end of each DVD-R.
11) The CD-Rs and DVD-Rs are printed with bar codes and basic information automatically from the database along with the case labels
12) The data is then exported as a batch file from RAPPER to the BBC's INFAX catalogue system.

Since the data is collected from various sources it inevitably contains errors and it is important that it is corrected before audio transfer. Examples of the discrepancies between catalogue entries and actual holdings are:-

- Duplicate entries - duplicate data imported into RAPPER from both 'Romeo' and 'Juliet' but in different order (ie track listings). Information produced to identify (duplicate) recordings found on DAT and 1/4 inch.
- Incomplete data - missing information relating to one or more songs in a session (especially difficult with the John Peel Sessions as each song would be on separate reels - ie. Where is the other reel? Has it been lost, destroyed or recorded over ?)
- Inconclusive data from 'Romeo' and/or 'Juliet' into RAPPER and information obtained from the tape box - Dates missing, vague abbreviations on box, identifying the variations found of the same programme e.g. Edited, Unedited, Live Mix, Studio Session, ROT, just speech inserts for full programme (edited, unedited), compilations.
- Wrong Tape in box - or right band or person but wrong category e.g. interviews not session.
- Tape catalogued then found to be blank when put through for transfer
- Additional unidentified recordings found on a tape - (In the early days tapes once left on desks were considered available for anyone needing one for an interview etc.)

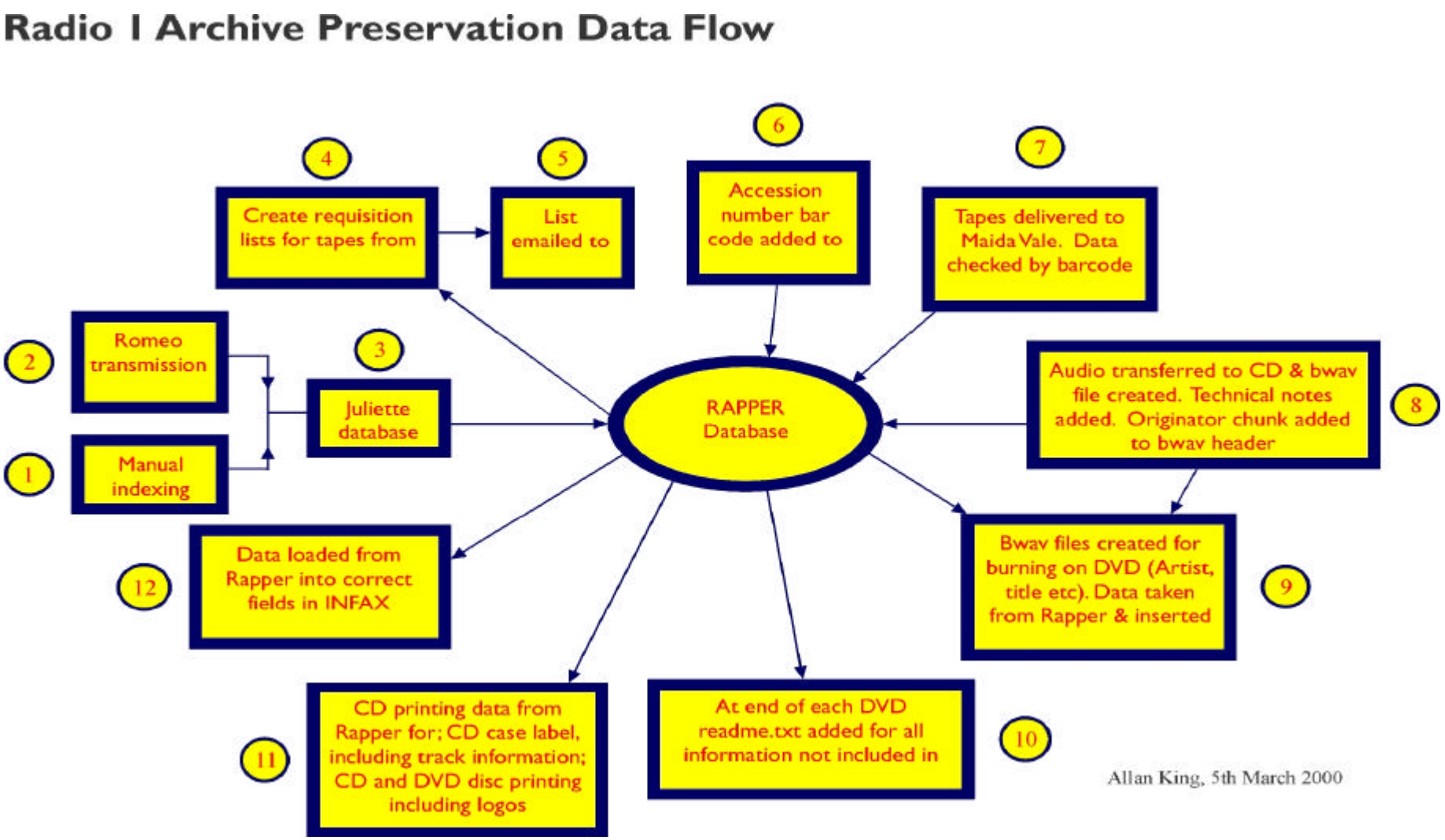

Figure 7: BBC Metadata Workflow

3.5.2. RAI: Audio Preservation

The following diagram is an overview of the RAI project. Greater detail will not be presented here, because for the metadata handling there are many similarities with the BBC work. The principal similarities are:

- barcoding at the point of entry to the system (unless material is already barcoded)
- use of a central control computer
- digitisation to a server rather than directly to the output media

The major difference is that the RAI process incorporates more quality control automation than is the case with the BBC process, allowing a higher throughput.

Both the BBC and the RAI processes produce audio CDs and file-format audio. RAI write the files on datatape (DLT), whereas the BBC uses DVD. RAI has a robot for 'near-line' access to the datatape. The BBC has not yet committed to mass storage, but the files on the DVD can be very efficiently transferred to whatever mass storage system the BBC will eventually acquire.

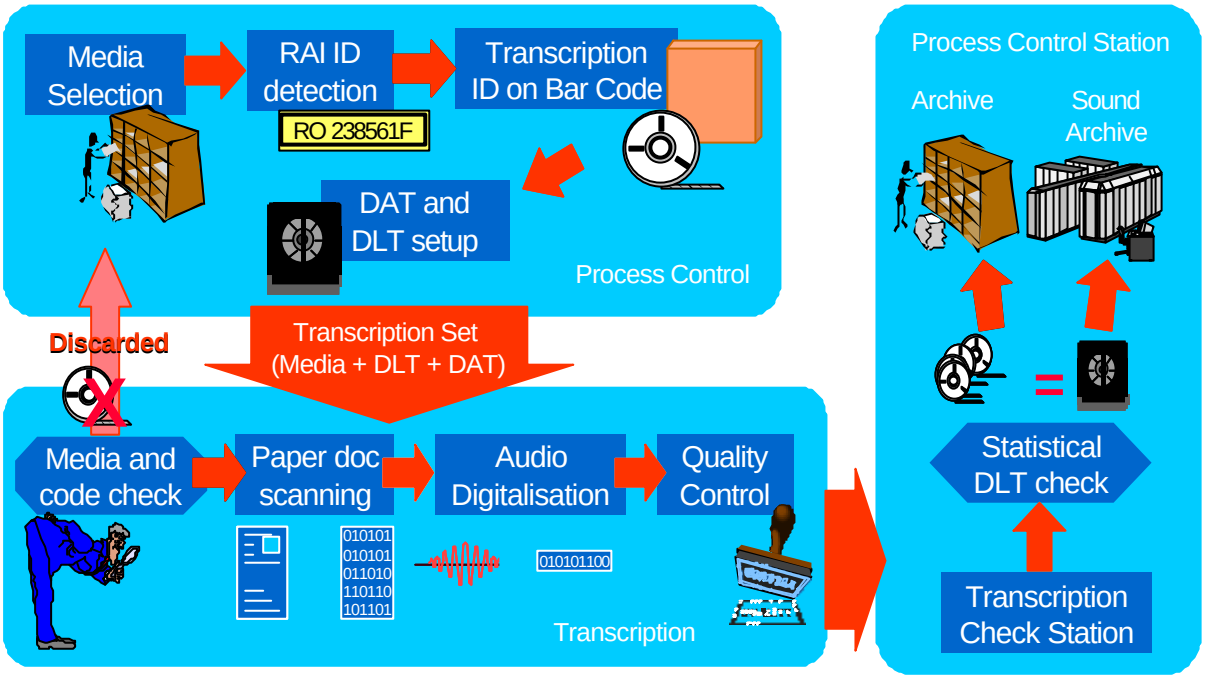

Figure 8: RAI Audio preservation Workflow

The volumes and media involved in the RAI project are given in the following table:

Item	Quantity	Average duration
¼" tapes	300.000	45'
CD	60.000	60'
Vinyl	60.000	35/40'
Total Items	420.000	
Total Hours		310.000

3.5.3. e-vod: Video Digitisation

The technical team of e-vod developed the BPS (Bank for Programs and Services), the first French video on demand service for schools, social and cultural centres. This bank is composed of 5.000 TV programs described in a web database. Each subscriber to the BPS service can order a programme on-line, and then receives it as an MPEG-2 file directly onto his PC's hard disk through a satellite link. At the same time, e-vod had to handle digitisation of new programmes and 30 hours of TV broadcast each night.

e-vod proposes a comprehensive and complete offer to catalogues owners, from digitising to secure e-commerce. This offer can be divided into the three following categories, along with the technologies they involve:

- Multi-level digitising and mass storage
- Secure distribution via all networks and media
- Catalogue management information system

The purpose of the e-vod work is not preservation, but digitisation. However the technical issues are virtually identical. The efficient solution, as for BBC and RAI, is automation with server-based storage and full computer control as the principal features, and with metadata handled automatically throughout the process. The following figure shows the architecture.

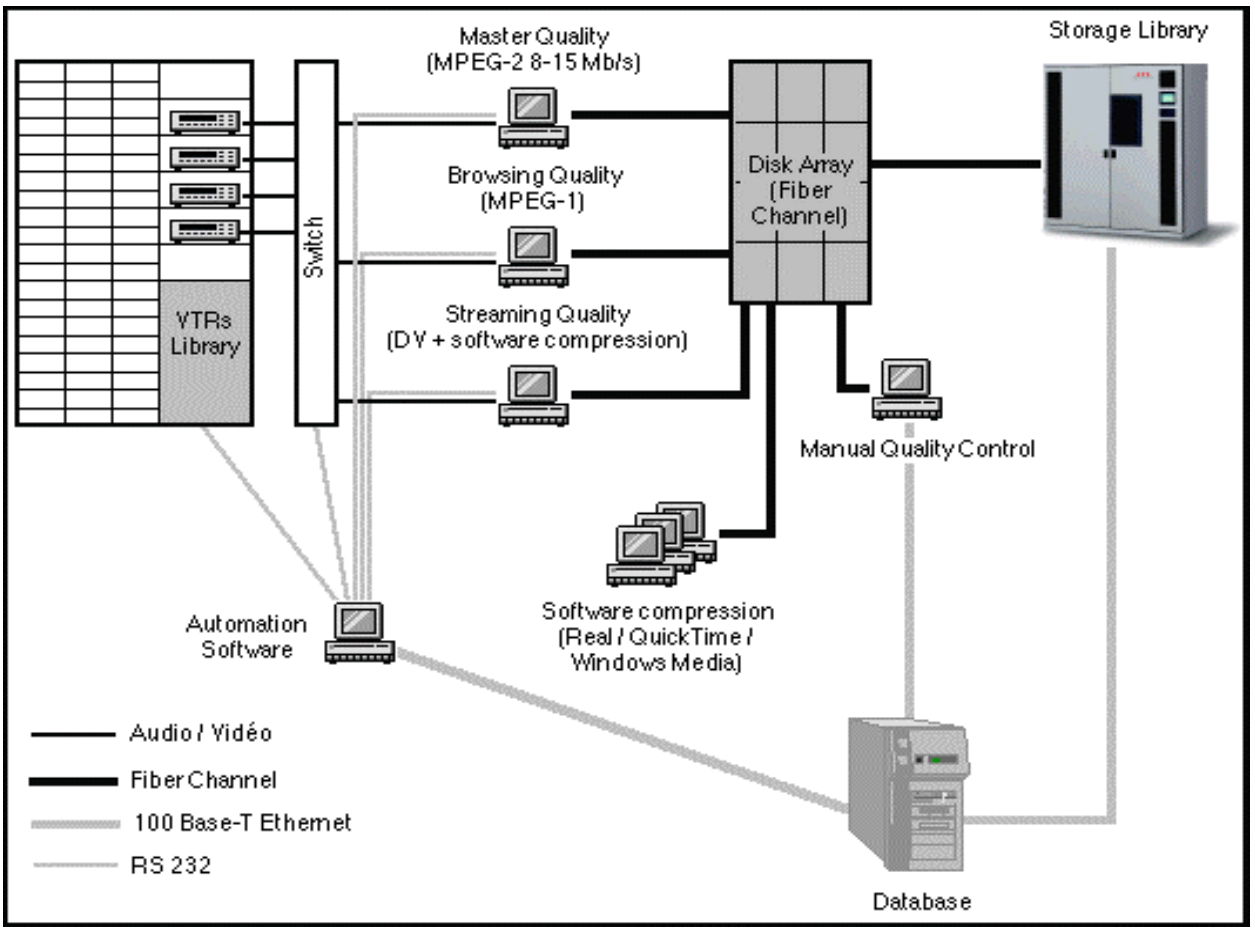

Figure 9: e-vod Digitisation Architecture

The key idea remains maximising throughput by using automation, which requires a server-based operation and process-control software.

4. Optimised model

4.1. PRESTO key links

The main finding of PRESTO is that automation and mass-production are the key to reducing the cost of preservation transfers. An associated finding is that automation of metadata handling is the general way to greater efficiency and lower costs. However there are numerous specific areas where technology can be improved: key links in the preservation chain. The key links worked on by PRESTO are listed in the following section.

4.1.1. Audio

- Audio Playback – Two aspects of audio playback technology are being addressed:

  For vinyl/shellac media (gramophone records), we are developing a double-arm turntable (based on readily-available commercial parts) with remote stop/start. This will allow rapid change of stylus or needle, including simultaneous use of both arms. At present copying from gramophone records is about twice as labour-intensive as working from audio tape.

  Sigma / Delta Evaluation: Sigma/Delta conversion is a relatively recent method of audio digitisation technology that could offer cost benefit improvements, but the technology has not been fully evaluated for the critical case of digitisation of archive material. The purpose of this work is to give archives the basic engineering information needed to know the benefits and risks of this approach.

- Audio Quality Control – The best judge of audio quality is the skilled human operator, but there are two areas where technology is also vital: in measuring technical parameters (such as stereo phase or tapedeck azimuth errors), and in monitoring multiple audio signals simultaneously. It is necessary for maximum efficiency to detect and allow correction of technical adjustment errors as soon as they occur, which implies a need for continuous monitoring, not just periodical equipment calibration checks. It is also necessary for maximum throughput for the human operator to run four or five or possibly more simultaneous transfers. The operator is then forced to 'spot check' the transfer process. Overall quality can be significantly reduced if the human checking is backed up by continuous automated checking for aspects such as overload, dropout, distortion and level. The outcome of the work will be a 'library' of software routines for signal quality analysis, logging and interpretation.

- Audio Lossless Compression – Algorithms exist for lossless compression of audio, which may be significant for allowing convenient and cost-effective use of existing media such as CD-ROM for audio storage. The available methods will be appraised for effectiveness, and suitability for use in the workflow of archive preservation transfers, and in the lifecycle of archive media.

4.1.2. Video

Manufacturers of videotape recorders (VTRs) cannot be expected to incorporate the advances in videotape technology into new players for old formats – because old formats are by definition obsolete. Three areas related to improving the performance of VTRs are under development, concentrating on 1" and ¾" (U-Matic) formats.

- Playback device improvement – investigating the available 'fixes' for using current knowledge to improve mechanical performance of old VTRs
- Digitisation quality monitoring / logging adding electronics based on microprocessor technology to detect and report upon playback errors. Such detection is built into modern VTRs, and allows the operator of a transfer to notice problems before they reach the stage where they cause visible defects.
- Time base corrector with drop out detection and compensation – adding a 'post-processing' module to VTR output, to use modern technology to stabilise the signal, correct for line dropouts, and provide the basis for more substantial correction (image restoration) of the material. Image restoration itself is a complex problem, and cost-effective solutions are being developed by the related project, BRAVA. http://www.ina.fr/recherche/projets/encours/brava/
- Multi-level encoding – current technology requires various codings of video material, for access in full quality, production quality, browse quality – or very low datarate access over the internet. PRESTO is developing a cost-effective workflow for producing these multiple encodings.

4.1.3. Film

- Auto-resplice – a major problem with film in television archives is old splices which fall apart when the film is played for transfer. Up to 80% of the budget of projects involving such film is spent on manually cleaning and re-splicing – and such film is very common for holding TV news stories of the 1950's right up to the end of the 1970's. PRESTO is developing a machine to automate the clean and resplice operations.
- Alternative film handling – flat-bed film scanning is being investigated for its potential for safer handling, and for speeding up the digitisation of film.
- Film Format Converter – there is no single standard format for digitised film, supporting efficient downconversion to all the broadcast standards required for broadcasting. PRESTO is developing a format, which could be a significant step in the transition to digital film.
- Lossless compression for film (and video) – lossy compression is common in broadcasting, but archives have a commitment to preservation in highest quality. Lossless compression allows storage savings of up to 75%, which could be worthwhile if the overall workflow could still be kept simple and reliable. The practical use of lossless compression for video and film, using existing algorithms, will be assessed.

4.1.4. Metadata management

- Common access to broadcast archives (Broadcast OPAC) – catalogue data from several PRESTO partners is being processed into a standard format, to make a form of 'union catalogue'. Such work is commonplace in book libraries and written archives, but broadcast archives rarely use standard library catalogues or standard access mechanisms such as the NISO standard Z39.50[2]. The point of the work is to demonstrate how such metadata conversion could be efficiently incorporated within preservation projects.

4.2. Server-based preservation

In addition to the key links just described, there are two general approaches to efficient projects for preservation transfers. One is the use of maximum possible automation of metadata, as discussed in 3. The second is the use of a server (electronic mass storage) as the buffer (central temporary storage mechanism) at the heart of the signal path for preservation transfers. The advantages of this approach are now presented:

Mass production is the key to saving money in broadcast archive preservation work. This statement is not entirely obvious. Many media archives, especially outside broadcasting, have collections that are far from uniform. The media type, quality, age, condition and recording standards can vary enormously from item to item, leading curators of such collections to completely reject the notion of mass transfers. Fortunately, broadcast archives tend to have large volumes of material on the same media, recorded to broadcast standard, and capable of being treated as uniform in its preservation requirements.

Mass production needs uniform material, and lots of it. It costs money to set up mass transfer: each operator needs to have as much hardware as he/she can operate, the whole area needs to have many operators to justify overall networking hardware and process control software, and shift work is needed to realise the full potential of the capital investment. This size of operation is only justified for volumes of at least 10,000 items.

Mass production efficiency is NOT just a lot of equipment and operators. There are two essential further items:

- A workflow based on a central server. Material (sound or images) is transferred from the old media onto the server as the first stage of the work. Thereafter material is held on the server for all further processing: labelling, cataloguing, any cleanup or restoration, production of various qualities of output (eg master, edit, broadcast, browse and internet versions of the audio/video).
- A computer programme to run the process. Material identification is scanned in from barcode at the beginning of the process, and thereafter each operation, each operator's daily work schedule, and each output is handled entirely automatically by the process controller. The data the computer programme is handling is basically the metadata associated with the digitised signals held on the server. This metadata automation programme becomes in effect the process controller, knowing where everything is, what has and hasn't happened, and issuing instructions to the operators to run the daily work, ensure no steps are ever missed, and finally producing and labelling the final outputs – the new media items – of the whole preservation project.

Maintaining Throughput: Mass production implies continuous workflow, which in turn implies that nothing can go wrong. In practice, especially dealing with aging material, things will always go wrong.

Two central issues in the process control are:

- Detecting when something is wrong
- Taking immediate action to "take the problem offline". An operator who is required to keep five workstations running simultaneously cannot let four stop because of dealing with a problem on one. Therefore the whole philosophy has to be to keep that operator, and those five (or however many) workstations, running continuously on non-problem material.

The A stream and the B stream: The continuous-throughput approach requires two streams: the A stream for the non-problem material, and the B stream to handle all the problems. The A stream has lowest achievable cost per item, and is kept up to speed by all possible means.

The B stream uses the traditional approach of letting a skilled person investigate the problem item and take individual corrective action. It needs to be emphasised that an archive that does not set up a factory essentially has **only** the B stream, and so has only highest possible rather than lowest possible costs per item.

One major problem with funding a preservation factory is that in many organisations a sort of B stream already exists. This leads to the situation where putting small amounts of material through this process – one form of so-called preservation on demand – can appear to be free, or nearly free. The equipment exists, the staff are there, transferring a few hundred items per year may not incur any additional visible cost (visible to an accountant). **The real cost of any B stream remains about double the cost (per item) of an A stream.** The real problem arises when an attempt is made to shift from hundreds of items per year to thousands. At that point, expanding the B stream capacity is the most ineffective way to approach preservation.

5. Conclusions

Detailed workflow will vary from project to project, but PRESTO has the following overall conclusions about workflow, metadata, quality and cost:

- Costs are reduced by minimising manual intervention
- Manual intervention is reduced by adding process-control automation

Process automation is of two sorts:

- Quality monitoring: hardware to monitor equipment operation, and software to monitor signal quality. Quality monitoring automation can increase throughput by a factor of five (five simultaneous transfers per operator instead of one)
- Metadata handling: bar-coding the first time media are handled, and use of a computer programme to track workflow thereafter

Finally, the way to organise preservation work to make it easiest to apply automation is with an architecture based around a central 'media server'. Media is digitised to the server as the first 'transfer', which allows maximum automation of all further processing. The server-based process also is the most effective way to take implement 'opportunity links' – those operations (such as making browse or internet versions) which are not essential to the narrow definition of preservation, but which open the most doors for further and future access.

References

[1] http://mpeg.telecomitalialab.com/standards/mpeg-7/mpeg-7.htm

[2] URL for NISO: http://www.niso.org/standards/standard_detail.cfm?std_id=465