International Association of Sound
and Audiovisual Archives

Internationale Vereinigung der
Schall- und audiovisuellen Archive

Association Internationale d'Archives
Sonores et Audiovisuelles

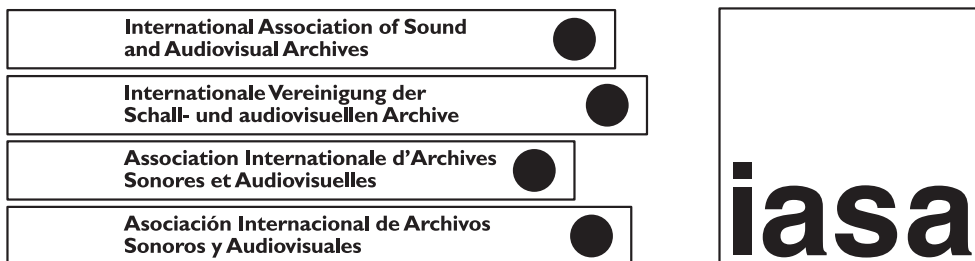Asociación Internacional de Archivos
Sonoros y Audiovisuales

iasa

Technical Committee
Standards, Recommended Practices, and Strategies

The Safeguarding of the Audiovisual Heritage: Ethics, Principles and Preservation Strategy

IASA-TC 03

International Association of Sound and Audiovisual Archives

Internationale Vereinigung der Schall- und audiovisuellen Archive

Association Internationale d'Archives Sonores et Audiovisuelles

Asociación Internacional de Archivos Sonoros y Audiovisuales

iasa

Technical Committee
Standards, Recommended Practices, and Strategies

The Safeguarding of the Audiovisual Heritage: Ethics, Principles and Preservation Strategy

IASA-TC 03

Co-Edited by Will Prentice and Lars Gaustad

Contributing authors:
Kevin Bradley, National Library of Australia, **Carl Fleischhauer**, Library of Congress,
Lars Gaustad, National Library of Norway and Chair of the IASA TC,
Bruce Gordon, Harvard University, **Will Prentice**, British Library,
Dietrich Schüller, Phonogrammarchiv, **Tommy Sjøberg**, Folkmusikens Hus

Reviewed by the membership of the IASA Technical Committee

THE PURPOSE OF THIS DOCUMENT

Through this document, the Technical Committee of the International Association of Sound and Audiovisual Archives seeks to inform the evolving challenge of safeguarding audiovisual heritage by offering these general principles and strategies for preservation. This advice identifies the tasks before us as archives, the nature of the objects for which we are responsible, the potential pitfalls and problem areas of preservation, and it guides the reader to focus on what is most important for their content, to enhance its survival in an unknowable future.

Our intent is to inform both those with financial responsibilities as well as those concerned with the more technical aspects of preservation, and so we hope that IASA-TC 03 will empower both to discover appropriate solutions that encompass those perspectives. More specific methods and technical details of audio preservation are well addressed in the publication IASA-TC 04: *Guidelines on the Production and Preservation of Digital Audio Objects*. For video preservation, the forthcoming IASA-TC 06: *Guidelines for the Preservation of Video Recordings* will serve the same purpose.

The future of preserving digital material for the long term will be one of managing a pathway between the choices we make now and those choices we must make in the future. We must act decisively now even though we know that technological developments will not necessarily align with those choices. Though no choice is a final one, a well informed decision will consider the process for navigating to the new.

Major changes in the current revision include a widening of scope to include moving image content, and a greater acknowledgement of the prevalence of file-based digital material alongside its carrier-based equivalent. While there are some significant changes to the language used in this version, the principles underlying the document are essentially unchanged.

0. ETHICAL CONSIDERATIONS

This document is not a Code of Ethics for all aspects of sound and audiovisual archiving. General ethical principles for sound and audiovisual archives (hereafter collectively referred to as audiovisual, unless otherwise stated) are covered in IASA Special Publication 06 *Ethical Principles for Sound and Audiovisual Archives*.

The guiding principles of this document can be summarised by the following statement:

> *Preservation enables us to disseminate to our successors as much of the information contained in our holdings as it is possible to achieve in our professional working environment. It is the responsibility of an archive to assess the needs of its current users and to anticipate to the extent possible the needs of future users, while balancing those needs against the condition of the archive and its contents.*

1. THE TASK OF AUDIOVISUAL ARCHIVES

A core responsibility of an archive is to ensure sustainable access to information. Essential to achieving this is the preservation of the information, which for audiovisual material requires the fulfilment of three related tasks:

1. The stability and optimal readability of the physical carrier bearing the information must be preserved, so far as is possible, through the use of best practices. This applies equally whether the information is held in analogue or digital form, file-based or otherwise.
2. The technological system required to access the information (replay equipment, spare parts, playback and format-migration software, expertise etc.) must itself be maintained or renewed, with sufficient capacity for the size of the collection.
3. Provision must be made to transfer the information to other sustainably accessible, file-based formats while access to the original information is still possible, ensuring that digitising or otherwise transcoding their holdings does not compromise the sonic and/or visual content, or other related information.

Comment:

Examples of what constitute best practices in audiovisual preservation can be found in IASA-TC 05: Handling and Storage of Audio and Video Carriers (2014), IASA-TC 04: Guidelines on the Production and Preservation of Digital Audio Objects (Second edition, 2009), and IASA-TC 06: Guidelines for the Preservation of Video Recordings (forthcoming).

The challenge of ensuring the sustainable accessibility of file-based formats through digital data management is at the heart of contemporary audiovisual archiving (see sections 12 and 13).

Technological advances can sometimes enable modern analogue replay equipment to retrieve more audio information from carriers than was possible at the time of recording. This is not currently the case with video, which is far more locked to original playback equipment. Modern techniques used in transferring analogue video may improve retrieval of the signal.

For a number of reasons, some of the holdings retained in, or offered to, audiovisual archives will not be the original recordings but copies. For the purposes of digitisation and preservation, these copies should be considered to be originals, unless earlier-generation or otherwise superior copies can be accessed through co-operation with other collection holders (see sections 6 and 16).

Although collection building and collection management per se are beyond the scope of this document, there are ethical and strategic aspects to the relationship between an archive and its potential contributors that should be addressed here. Technological change increasingly democratises the creation of sound and audiovisual content, and increases the number of formats in which material is created. Much of this material may at some stage find a legitimate home within an archive, and for reasons discussed below, the format in which content is created or submitted can have a significant effect on the subsequent use of that material and on its preservation. It is important therefore to raise awareness among potential contributors to archives—whether professional producers or the general public—regarding the consequences of using data reduction, proprietary codecs or other content-compromising schemes described in sections 10 and 11.

2.   PRIMARY AND SECONDARY INFORMATION

Any archival document consists of multiple forms of information. Some may be considered primary information. This includes the obviously audible or visible time-based content, i.e., the sound or picture signals. Other forms of information may be considered secondary, where they play a contextual or supporting role to the primary information. For example this might include information about the contents (perhaps written on a physical carrier), information about the carrier itself, or for video, timecode embedded within the video stream.

Both primary and secondary information form part of an audiovisual document, whether carrier-based or file-based. The relative importance of the two will vary depending on the content, the type of carrier and the needs of users, both present and future. Secondary information, however, becomes a crucial factor in the authentication of primary information transferred from another carrier format, or as a potential source for other analyses or research. Secondary information may be present in file-based born digital content or in content on physical carriers. When file-based content is format-migrated, or when carrier-based content is transferred to file-based formats, care must be taken to retain the secondary information. A minimum combination of primary and secondary information is required to preserve a document's essence sustainably, and it is the responsibility of the archive to define that combination of information explicitly, through careful analysis of actual and potential use as well as ethical, legal or other institutionally mandated considerations.

Comment:

Any and all metadata may be considered secondary information, including machine-actionable data allowing specific functionality, such as DVD menus or video game functionality.

In discussing video, the terms ancillary and associated data are often used, to describe such things as timecode, captions, and any other information that is not strictly sound or picture.

Physical carriers can often be considered as significant cultural objects in their own right, for example mass-produced sound discs, and the scholarly and cultural value of commercial disc sleeves and labels should be borne in mind.

Researchers may find the secondary information held in video timecodes of special interest as they can provide clues to a television producer's editing processes.

When digitising film, for reasons of authenticity it is important to digitise all information written or recorded on the film prior to and after the essential images, including the geometrics of the sprockets, either as part of the preservation copy, or at least in the metadata.

3. THE INSTABILITY AND VULNERABILITY OF AUDIOVISUAL CARRIERS

For traditional paper and film-based documents, the long-term preservation of the original carrier is, with a few exceptions, generally feasible. Printed or handwritten text as well as film-based documents may remain fully human-readable even when damaged, whereas the continuous, time-based nature of audiovisual documents means that any compromise in the integrity of the document will result in the loss of information.

In addition, audiovisual carriers are generally more vulnerable than conventional text documents to damage caused by poor handling, poorly maintained equipment or by poor storage. Many audiovisual carriers, especially magnetic recordings, laminated instantaneous discs, and nitrate film, have relatively short life expectancies due to their physical composition. Whereas script has a high level of redundancy, which keeps text documents often readable even in damaged condition, audiovisual documents are representations of physical facts or processes: their redundancy is low, as each detail is potential information that must be preserved, which calls for highest possible integrity standards.

These factors have led to the development of a wide range of best practices for the storage and cleaning of carriers, and for the transfer of content to digital file-based formats. Passive preservation is discussed in detail in IASA-TC 05: *Handling and Storage of Audio and Video Carriers*.

Due to the high density of information, digital carriers are generally more vulnerable to loss of information through damage than analogue carriers. Life expectancy concerns particularly arise in the case of the storage media used in most computer-based storage and data management systems. Their useful life is generally short—from three to ten years—due to a combination of system and storage-media-format obsolescence, as well as risks related to the high density of the data carried by data-storage media.

Comment:

High data density and the risk of data loss is a particular concern for digital video carriers containing Metal Evaporated (ME) tape.

The level of risk to a carrier will depend in part upon its vulnerability to decay or damage. It also depends upon the storage conditions under which the carrier has been held, upon the quality and maintenance of replay equipment, and upon the professional skills of the operator.

Digital carriers will fail without warning, and without the audible or visible clues of gradual degradation that analogue carriers present. Damage to a digital carrier's logical structure may also render the content inaccessible.

4. OBSOLESCENCE OF FORMATS

No format, whether carrier-based or file-based, will be playable forever, and for some the end is in sight. Since the 1990s there has been a clear shift away from carrier-based formats that store content in a manner specific to a particular physical medium, towards file formats, storing content as data in a computer environment. This market-driven obsolescence of formats obliges us to acknowledge that the window of opportunity for digitally preserving carrier-based content is finite. Beyond a certain point, the maintenance of obsolete replay systems will become unaffordable, and so our access to content held on certain media will cease.

The actual time available in which to digitally reformat carrier-based content may be extended somewhat, through careful stockpiling of equipment, spare parts, service manuals and other accessories that have been or shortly will be discontinued, and through the careful retention of maintenance and operational skills. It will also vary according to format, but ultimately it is finite. As of 2016, it is widely accepted within the global audiovisual archival community that we have between 10 and 15 years in which to digitally preserve all carrier-based audiovisual content held on magnetic media. Indeed, for some magnetic-based formats such as MII video, retrieval of the content is already practically impossible. Therefore, by around 2030, digitisation of even the most common magnetic media may be beyond the reach of most archives. For other formats the timescale may be longer or shorter.

For file-based content, obsolescence of the physical carrier format, the operating system or specific coding or file wrapper formats may endanger either the bits themselves or the ability to meaningfully interpret the bits. However, the fact that these problems are common to all modern industry makes them easier to manage than individual legacy audiovisual formats driven by specialist consumer markets.

5.   SAFEGUARDING THE INFORMATION

a. By preservation of the carrier

Although the life of audiovisual carriers cannot be extended indefinitely, efforts must be made to preserve carriers in useable condition for as long as is feasible.

For carrier-based content, preservation of the carrier requires storage in an environment suited to the purpose, separating primary and secondary information sources where appropriate, and performing routine maintenance and cleaning as needed. Maintenance includes the regular checking of reference signals and frames, if available, on analogue carriers, and regular checking of the data integrity of digital carriers. In addition, the equipment used for handling and replay must meet the physical requirements of the carriers. Preservation includes minimising the use of original copies by making access copies available.

b. By subsequent copying of the information

Because the life expectancy of carriers and the availability of hardware are limited, the preservation of the document in the long term can only be achieved by copying the contents to new carriers/systems while this remains possible.

In the analogue domain, primary information suffers an increase in degradation each time it is copied. Only the digital domain offers the possibility, when done diligently, of lossless copying when refreshing or migrating recordings (see section 12). For the long-term preservation of primary information held on an analogue carrier it is necessary, therefore, to first transfer it to the digital domain.

Separating the primary information from the original carrier raises the question of future authentication of the sound and images. Future users may only have access to an audiovisual document in the form of a file-based copy; in this event, the importance of adequate secondary or contextual information increases. Secondary information held visually on tape boxes, disc sleeves or labels may be best captured and retained in associated image files, in which case this information should be documented and reproduced according to recognised archival standards for generating such content. Other types of secondary information such as descriptions of the original format may take the form of preservation metadata, and so should be recorded in a systematic manner, and be made accessible together with the primary information (see section 14). By this means, future users can be better assured of the authenticity of the document.

6. SELECTION OF BEST COPY AND CARRIER PREPARATION

An archive may possess multiple copies of certain content, perhaps including commercial audio recordings or motion picture films. Where holdings include more than one copy, the best example should be selected before reformatting its content. In the case of magnetically recorded or file-based sound or video recordings, an archive or production house may hold multiple versions for different purposes, e.g., a master version and a copy made for some form of distribution. Here again, care should be taken to select the most appropriate copy prior to reformatting. Additionally, cautious and appropriate cleaning and restoration procedures may be necessary to optimise signal retrieval and reproduction.

Comment:

With mass replicated carriers—specifically mechanical and optical carriers—the replay quality of different copies may vary considerably because of the way that they have previously been handled and stored. It may therefore be advantageous to extend the search for the best copies to other collections on a national or even international scale (see section 16).

Non-replicated recordings may often be available in two or more versions. For example, a recording held on its original carrier (tape, cassette, film etc.) may also have been copied to an archival master. While the archival master (or subsequent generation copy in general) may often be in better physical condition, it may be of inferior signal quality due to poorer earlier transfer technology and the unavoidable signal degradation caused by the analogue copying process. Consequently, the signal quality of the various available copies must be compared.

Appropriate cleaning and restoration procedures can significantly improve signal retrieval. However, utmost care must be taken to balance any potential benefits against the risk of further deterioration, caused by subjecting fragile or deteriorating carriers to such procedures. It is good practice to minimise the handling of carriers at all times.

Some carriers may be so heavily deteriorated that even an attempt to replay them might place their content at risk. In such cases, a careful and informed evaluation needs to be made between the likelihood of damage through replay immediately, or through delaying replay until a less risky replay technology becomes available.

7. OPTIMAL SIGNAL RETRIEVAL FROM ORIGINAL CARRIERS

Optimal retrieval of an analogue signal, where replay distortions are kept to an absolute minimum, can only be achieved by modern, well maintained replay equipment, ideally of the latest generation. When replaying historical formats, replay parameters (such as speed, playback equalisation, track format, type of time base stabilisation, etc.) must be chosen objectively, and be based upon knowledge of the given historical format.

Certain adjustments to replay equipment may be necessary, in order to align with original recording characteristics and to optimize the retrieval of the recorded signal. For example, azimuth error is common in analogue magnetic tape recordings, and can only be corrected during the replay of the original carrier at the time of digitisation. Similarly, storage-related print-through must be minimised at the point of signal extraction. Other minute inaccuracies in the tape path adjustment of original recordings can also cause a considerable and avoidable rise in errors.

For video, certain types of dropouts are best compensated for at transfer time. Where motion picture film is being copied, some scratches can best be eliminated or suppressed by the use of liquid bath in the film printer when the transfer takes place. In a digital scanning transfer, the use of specialised diffuse light sources can have the same effect.

In order to minimise possible damage to the original carriers, replay equipment must be regularly maintained to professional standards. To aid in this and to diagnose emerging problems, calibration media suitable for the replay equipment must be used whenever obtainable.

For digital carrier-based formats, different players or readers may retrieve data from the same carrier in varying ways, not all of which will successfully present the bitstream for transfer. In order to evaluate and detect such problems, error monitoring during real-time replay, or error reporting after high-speed ripping, is imperative. The presence of uncorrectable errors copied to resultant files for preservation must be documented.

Digital carrier-based formats may contain various types of sub-code information, that is, secondary information written in parallel with the primary information bitstream. Incompatibilities between recording and replay devices can result in this information being retrieved incorrectly or not at all. Understanding the properties of a given format or collection, including any sub-code information, and defining the minimum required combination of primary and secondary information prior to its digitisation, is of utmost importance (see section 2).

It is not always an easy task to assess the correct replay parameters for a given analogue audiovisual document if objective information on the recording format parameters is missing. As in other fields of historical research, the use of cautiously chosen approximations is permissible when necessary. As a matter of principle, however, all such decisions must be documented, and irreversible steps should be avoided. All unnecessary subjective treatments must only be applied to access copies.

Comment:

Inadequate signal retrieval from original documents is very often the result of a lack of professional knowledge, or the use of inappropriate equipment. It is hard to overstate the importance of operator skill and experience, as well as the availability of specialised equipment, when reformatting challenging materials. Optical sound tracks for film-based motion pictures, for example, can be very challenging to transfer, and the role of highly specialised equipment can be crucial.

In some circumstances it may be appropriate to take a multi-layered approach to choosing replay parameters. This might involve digitisation and creation of master preservation files without playback equalisation, and applying equalisation either in the creation of access files, or as a software process at the time of access.

For example where one-light transfers from film are deemed appropriate, the RGB output should be adjusted to get the maximum colour information from each channel, to correct for colour fading without introducing any clipping.

Best practices for the transfer of motion picture film for preservation are still in development, with some cutting edge work advancing under the auspices of the Academy of Motion Picture Arts and Sciences (AMPAS) and the Society of Motion Picture and Television Engineers (SMPTE). These new developments will help standardise approaches that will have a special impact on the capture of colour and the representation of tonal variation in the original film. However, systems that implement these new developments are not yet widely available and the approach is not yet employed in memory institutions.

The systematic retrieval of sub-code information from digital carrier-based formats, as a method of safeguarding useful secondary information, is still a widely neglected subject. This is largely due to incompatibilities between the sub-code formats of different players and interfaces. As yet, few if any standards have been widely adopted for the further retention of this information in file-based formats. Compatibility problems can also often be encountered in the replay of recordable or rewritable optical disks.

The principles described in this section apply unambiguously where primary information exists in the form of documentary records, whether documenting artistic performance or other forms of actuality. Where the primary information exists as part of an art object however, for example where sculpture or installation art has an audiovisual component, there may be an additional ethical requirement to preserve original reproduction distortions, and therefore diverge from these principles, in order to honour the intentions of the artist. Determining the intentions of the original creator(s) may be necessary in choosing how such art objects may best be represented in a file-based environment.

8. UNMODIFIED TRANSFER TO A NEW TARGET FORMAT

It is mandatory that transfers made from old to new archive formats be carried out with the intention of producing the closest possible surrogate. Above all, subjective alterations or "improvements", such as de-noising or de-graining of film, must be avoided. Subjective alterations effectively rewrite the historical document according to the perspective of the operator undertaking the change, thus undermining the most basic principles of preservation.

The signal that the original recording engineer intended to capture is only part of a given audiovisual document. Unintended and undesirable artefacts (e.g., noise, distortions, drop-outs) are also part of it, whether caused by the limitations of historical recording technology, or subsequently added to the original signal by general use, mishandling or poor storage.

In some instances, apparent "imperfections" in a recording may be objectively corrected at the time of digitisation, by adjusting replay parameters to optimally retrieve the intended signal (see section 7). Generally speaking however, both the signal and the artefacts should be preserved with the utmost accuracy. It is essential that the full dynamic range, frequency response and/or image resolution of the original are transferred.

The careful documentation of all parameters chosen and procedures employed in the transfer process likewise is essential.

Comment:

Alterations in transferring from old to new formats are unavoidable in some circumstances, for example when converting an analogue composite video signal to a digital colour-difference bitstream.

9. PRESERVATION OF CARRIERS AND REPLAY SYSTEMS AFTER TRANSFER

In the future, technological developments may allow improved information retrieval from physical audiovisual carriers. Similarly, new research findings or methodologies may allow users to identify additional secondary information in the original carriers.

Because of this potential for improvements in information retrieval, transfers of primary and secondary information from carrier-based formats cannot necessarily be considered definitive. Original physical carriers and suitable reproduction equipment must therefore be preserved after digitisation of their contents whenever possible.

It is quite possible however, that carrier degradation, technological obsolescence and the sheer cost of the digitisation process will prevent any further attempt. All transfers must therefore be carried out to the highest standards possible at that time.

Comment:

Original analogue sound carriers may contain secondary information which falls outside the frequency range of the primary information, and which may assist in correcting inaccuracies in the original recording. Most current transfer technologies result in an irretrievable loss of this information. For analogue magnetic audio tape for example, information about speed fluctuations (wow and flutter) may be found in the variations of the reproduced bias frequency, AC traces or background noise. Processes that can use this information to correct the primary information are now available and may become part of future transfer routines.

Another recent improvement in audio transfer technology is the non-contact, optical scanning of primary content from mechanical sound carriers. Best practices have yet to be developed, however see section 10 comments.

10. DIGITAL TARGET FORMATS AND ACCURACY

As with all forms of digital technology, digital coding schemes are subject to ongoing development. As such, discussion around the most appropriate formats for preservation will also continue to evolve. Irrespective of the options available however, several principles can be applied in choosing target formats.

- File-based formats offer greater data security and integrity monitoring capability than do carrier-based formats containing data streams such as DAT, audio CD or Digital Betacam.

- When transferring digital carrier-based content (for example from DAT or DV cassette formats) the resultant file must, when deemed appropriate, retain the coding scheme of the original data stream. Where this is not appropriate, for example where a lossy and proprietary coding scheme has been used (see section 11), a coding scheme should be chosen which preserves the integrity of the original.

- An essential requirement of any archival file format is that coding schemes used for preservation purposes be openly defined, and not proprietary to a limited number of manufacturers.

- Where there is little or no consensus throughout the archival community on the choice of target format for a given purpose, a repository must choose a format for which they can be at least relatively confident of their own ability to support it sustainably. This would require sufficient available resource including expertise, as well as ongoing wider industry support for the format.

- A repository must ensure that a chosen target format will retain the minimum required combination of primary and secondary information.

Comment:

Preservation master recordings are generally carried by a target format that consists of a single file, in which a container (wrapper) carries the primary sound or sound-and-picture information, together with secondary information like captions, subtitles, timecode, and other ancillary data. In some cases, however, the secondary information may be carried in what are sometimes called "sidecar" files. This approach is not uncommon for subtitles or captions, and may be used for such corollary materials as record labels.

For audio, the Broadcast WAVE (BWF) format has become a de-facto standard. This format is officially recommended by the Technical Committee (see IASA-TC 04, 6.1.2.1). Broadcast WAVE files, like all WAVE files, cannot exceed 4GB in size, and are limited to mono or two-channel stereo recordings. To accommodate greater amounts of audio data and multiple audio channels, the European Broadcast Union has defined the RF64 BWF file, with a maximum file size of approximately 16 exabytes and up to 18 channels.

For digitisation of original analogue audio recordings, IASA recommends a minimum digital resolution of 48 kHz sampling rate at 24 bit word length, using linear pulse code modulation (LPCM) encoding. In heritage/memory institutions a resolution of 96 kHz / 24 bit has become widely adopted. Better transfers of the unintended parts of a sound document now (see section 8) will make the future removal of these artefacts by digital signal processing easier when making access copies. Because of the transient character of consonants, speech recordings must be treated like music recordings.

When primary information on disc and cylinder sound recordings are captured by non-contact optical scanning techniques, the scanning data itself may comprise the main element in the preservation master file, rather than a subsequently derived conventional audio bitstream.

In memory institutions, target formats for moving image preservation masters are in the early phases of implementation. For video, several institutions have been using a variant of the MXF wrapper standardised by SMPTE, with the picture signal encoded as lossless-compressed JPEG 2000. Meanwhile, other institutions are moving forward with the FFV1 lossless encoding, carrying the picture signal and accompanying soundtracks in wrappers such as QuickTime, Matroska, or AVI.

The most frequently selected target format for memory institution motion picture film scanning is DPX, standardised by SMPTE. At the same time, some archives are exploring approaches that will permit the carriage of synchronised sound and picture signals in the same wrapper, and/or the ability to incorporate additional colour and tonal data. These explorations entail the reformatting of the initially captured DPX picture signals (and soundtracks) into preservation master formats like those selected for video, e.g., lossless JPEG 2000 in MXF or FFV1 picture in QuickTime or Matroska.

In some circumstances it may not practically be possible to migrate audiovisual content. This could be due to specific integral functionality as encountered in video games for example, or the use of copy protection technology. Future access (and thus preservation) may therefore depend on the emulation of the original operating systems and/or application software.

Archives may acquire material in file-based forms, whose transcoding to archival formats may result in irreversible changes being made to the representation of the content. In such cases, authenticity and the promise of better transcoding methods in the future must be considered. The archive may choose to retain the original (as-acquired) file, as well as the transcoded version that is considered a better bet for long-term playability, or simply to transcode, retain the new copies, and delete the originals. The latter option may apply particularly in "edge" cases such as video clips that have been gathered as a part of a Web harvesting project.

In the very long term, further migration from any given format would seem inevitable. Therefore, as far as is possible, a repository must aim to ensure that future migration from any chosen target format will equally preserve this information.

11. DATA COMPRESSION AND DATA REDUCTION

For long-term preservation purposes, target formats employing data reduction (often called, incorrectly, data "compression") should not be used when encoding from original analogue or linear digital recordings. Such so-called "lossy codecs" based on perceptual coding result in the irretrievable loss of parts of the primary information. The results of such data reduction may sound and look identical or very similar to the unreduced linear signal, but further use of the data-reduced signal will be much more likely to result in degradation of the primary content.

While there is no objection in principle to the use of lossless (fully reversible) compression, any resultant saving in storage costs would need to be offset against the increased risk that the tools required to decode the files might be unavailable or insufficiently supported in the future. Both lossy and lossless data compression schemes produce data streams that are more susceptible to minor read errors than are linear encoded streams, and therefore the content of those compressed streams is more likely to be corrupted to a much larger extent by those errors than a linear encoded stream would be.

This archival principle should also be applied, whenever possible, to the creation of original recordings made with the intention of being archived. However, if content comes to an archive having been recorded in a data reduced, non-linear format however, it must be preserved faithfully as-is.

Comment:

Data reduction is a powerful tool in the dissemination of audiovisual content. Its use for preservation, however, is counter to the ethical principle of preserving as much of the primary information as possible. Data reduction does not permit the restoration of the signal to its original condition and will, in addition, limit the further use of the recording because of the artefacts generated when cascading perceptually coded material, for example, in the making of a new programme incorporating the original sounds and images.

Because of the sheer amount of data required to store digital video signals, the use of data reduction for production formats has been and remains widespread. Ideally, non-linear encoding formats should be preserved in their original form. A major problem may arise however, when the format of origination is of a proprietary character such as the MiniDisc and DVCAM (see IASA-TC 04, 5.5.12.1; IASA-TC 06). The primary information of such recordings may be logically migrated to a preservation supported format, or the coding may be kept as-is. This will often be the decision of the archives digital preservation policies.

12. DATA MANAGEMENT: ARCHIVING PRINCIPLES IN A FILE-BASED ENVIRONMENT

The core actions in file-based archiving pertain to *bit preservation*, i.e., a set of actions that maintain the integrity of the digital data ("bitstreams") that are being managed by the responsible institution.

Actions beyond bit preservation will ultimately be needed when the formatting of the content is obsolescent. The most common action will be format migration, although (as noted in section 10 comments) there may be contexts in which system emulation is required. While bit preservation decisions may be left to information technology specialists and appropriate software and hardware applications, the actions beyond bit preservation will benefit from the involvement of people with curatorial responsibilities. What is at stake requires consideration of the significant properties of the content, the makeup of the research community being served, and an assessment of format obsolescence and the options for the new target formats.

Data management must observe the following core principles:

- Files are generally placed in storage systems by copying. This process must produce duplicates that are *verifiably identical* to the originals. This process of data integrity checking can be achieved through the prior creation of a checksum, also known as a hash or digest. The process of verification should take place immediately after the creation of the copy, ideally as an automated procedure.

- The ongoing data integrity of file-based content must be checked at regular intervals to ensure that it can be read exactly as it was written, with no errors or changes.

- Depending on the original file format however, it may be desirable to transcode to a new target format rather than simply copy from the original file (see sections 10 & 11). This process is known as *format migration*.

- Digital content, whether file- or carrier-based, must be copied to a new physical carrier before uncorrectable errors occur. When the original and target formats are the same, this process is known as *refreshment* or *media migration*.

- It is essential to keep *at least* two digital preservation copies, ideally more, and to use further dedicated copies for access as appropriate. The preservation copies should be kept in different geographic locations whenever possible. Additional security may also be provided by the use of different storage technologies for each set of preservation copies. When choosing which technologies to use, it should be borne in mind that a strategy will only be as strong as its weakest link.

- Access copies should be made whenever possible. Unlike archival master files however, such access or distribution copies may be subjectively modified, depending on the requirements of users. Data reduction may also be employed when compatible with user requirements. As with the creation of archival masters, careful documentation of all parameters and procedures employed is essential.

- Where possible, checks to ensure data integrity should be automated, as is possible with equipment within trusted digital repositories. If this is not possible, then manual checks will need to be undertaken, on a statistically significant basis.

Comment:

While these principles apply equally to any form of file-based preservation, the relatively large file sizes and time-based nature of audiovisual content demand that storage and bandwidth capacities be considered carefully.

Essentially, these principles are the same as those recommended for the analogue world. One fundamental difference, however, is the qualitative dimension of the file-based digital world, which permits objective validation of the integrity of recordings. Regular data integrity monitoring is amongst the core obligations of digital preservation routines. Digital carriers and systems can and do fail, without warning, at any time. Strategies for minimising risks to digital archives are greatly supported by networking between the primary collection, the user and backup archives.

13. STORAGE OF FILES FOR THE LONG TERM

Infrastructure permitting automated checking of data integrity, media migration, and, finally, format migration with a minimum use of manpower is now in regular use within the archival community (see IASA-TC 04, 6.2). At the time of writing, these range from small-scale systems of around 16 TB to petabyte size repositories. Such management software exists both as expensive proprietary solutions and as freeware. Even if an archive decides such a system is beyond its means, it should not postpone digitisation, but consider using discrete data carriers such as data tapes or HDDs for offline storage, and initiate manual checking procedures.

Responsible preservation of digital data requires systems and a technical infrastructure, the monitoring of the condition of files, and the existence of plans for media migration and format migration. All of these topics and more are discussed in standards associated with the Open Archival Information System (OAIS) Reference Model (ISO 14721) and in documents pertaining to Trusted Digital Repositories (ISO 16363).

Comment:

While the cost of hardware and software for long term preservation is within reach of many audiovisual archives, the archive must ensure that it possesses the knowledge base required to run and maintain such a system. In manual approaches, as opposed to automated storage systems, the lower cost of hardware and software has to be offset against a considerably increased requirement for labour, with all its implications in terms of risk to the carriers and staff costs (see IASA-TC 04, 6.5).

14. PRESERVATION METADATA

In its broadest sense, preservation metadata could include any contextual information required to provide sustainable access to content. In addition to technical requirements, this might include information required to authenticate the content for example. In this broad sense then, preservation metadata should contain full details about:

- any non-file-based carriers the content has been held on, including their condition

- the replay equipment used in the transfer process, and its parameters

- the capture equipment used, including known rendering software

- format information on the resultant file, including the digital resolution

- the operators involved in the process

- checksum – the digital signature that permits authentication of the file

- details of any secondary information sources.

In practice, metadata is often separated into categories including descriptive, administrative structural and preservation metadata. Preservation metadata in this specific sense is mandatory to evaluate the technical parameters of a recording, and to draw appropriate conclusions for the management of preservation. A subset of preservation metadata, namely the metadata necessary to faithfully render the primary information, may be considered an indispensable part of an AV document.

It is strongly recommended that metadata be written according to established standards, in as consistent a fashion as possible. Writing metadata in a machine-actionable form (for example using XML schemas) has the further significant advantage of enabling automation of certain preservation and dissemination actions.

Comment:

Metadata, often described as "data about data" is, in the digital environment, a detailed and specific extension of cataloguing practice. However, when associated with digital collections, it is a necessary part of their use and control. A Preservation Metadata Set is a statement of the information required to manage preservation of digital collections. Preservation metadata will be a key component in the preservation and management of any digital collection and must be designed to support future preservation strategies. A vital component of preservation metadata is the checksum or digest of a file, which is essential in monitoring data integrity and verifying authenticity. As such, it may be compared to the fingerprint of a given file.

The most thorough articulation of preservation metadata is represented by PREMIS (http://www.loc.gov/standards/premis/), the product of an international working group active from 2003–2005, and subsequently updated and revised by members of the digital library community. PREMIS is conceptualised around four categories: the Object, Event, Agent and Rights.

The Object entity pertains to what is stored and managed in the preservation repository.

The Event entity aggregates information about actions that affect objects in the repository, vital for maintaining the digital provenance of an object, which in turn is important in demonstrating the authenticity of the object.

Agents are actors that have roles in events and in rights statements and they can be people, organisations, or software applications.

Issues pertaining to rights or other restrictions arise not only when providing access to content but also when preserving it, since most preservation strategies involve making identical copies and derivative versions of digital objects, actions that may be limited by copyright law or by other restrictions, e.g., requirements imposed by donors. PREMIS rights metadata aggregates information about restrictions that are directly relevant to preserving objects in the repository.

Metadata can be stored with the resource it describes (e.g., within file formats that support descriptive headers or file wrappers), separate from the resource (e.g., within an external catalogue) or separate but linked to the resource (e.g., a file linked to the digital object in a repository structure). Each strategy has particular benefits and disadvantages. It is possible, and probably desirable, to use these strategies in parallel. The use of standardised wrappers is emerging as a trend in digital preservation of audiovisual material, because of their ability to handle file relationships. Wrappers also allow the possibility of retaining all of a file's primary information within the digital object.

15. PRIORITISATION

Sooner or later, all sound and audiovisual content destined for long-term preservation will have to be transferred to file-based digital storage repositories. As the transfer process is time consuming and cost intensive, it should follow a strategy based on the individual situation of the collection and the specific policy of an archive. Generally, priority should be given to those documents that are at greatest risk, through either degradation or technical obsolescence (see sections 3 and 4).

Carriers likely to degrade due to inherent instability, age or improper handling may include:

- wax or celluloid cylinders
- nitrate film
- instantaneous audio discs of all types, especially "lacquer" discs
- acetate tapes
- acetate film showing signs of colour fading, unless stored frozen
- ½" EIAJ video tapes
- U-matic tapes
- recordable optical media (CD-R, DVD-R etc.)

Prioritisation must be seen within the wider picture of technological obsolescence however. Many common carrier-based formats, although degrading, will outlast our ability to replay them, and this applies in particular to most magnetic tape-based formats (see section 4). For many, perhaps most archives, obsolescence will pose a more immediate threat to collections than degradation.

Where an archive intends to digitise their audiovisual collection themselves, they are strongly advised to check the quantity and quality of their equipment against the size of their holdings, and to take immediate action to ensure they have sufficient modern equipment and supporting infrastructure to enable the optimal replay of their entire holdings (see section 7).

Comment:

With one exception, the above list of carriers does not imply an order of priority. Prioritisation within each collection must be based on examination, and will depend on the individual rates of decay of the carriers, the availability of suitable playback equipment, and, to a lesser extent, the existence of duplicate copies of the material.

The exception is that priority must be given to "lacquer" or "acetate" discs. Even when these discs are playable they are at grave risk of suddenly cracking or crazing without warning. This is because of the steadily increasing stress between the lacquer coating and the supporting base plate. This stress is generated by shrinkage of the lacquer coating. Lacquer discs should, therefore, be given the highest priority in the copying program.

Format obsolescence is also associated with a vanishing market for test (calibration) equipment including test tapes, discs and cassettes, as well as ancillary accessories such as empty reels, cassette housings, splicing and leader tapes, etc. Test material is still supported by a few vendors for some sound and film formats.

16. CO-OPERATION

The exchange of information between archives performing preservation work is an ethical obligation. National and international co-operation in this respect is imperative, in particular in the dissemination of information to smaller or less-specialised collections for whom carrying out all the necessary stages of digital preservation is not feasible due to lack of resources.

Comment:

The greater part of the world's heritage of audiovisual documents reflecting the linguistic and cultural diversity of mankind is kept by comparatively small institutions, by scholars and other private individuals. Co-operation and exchange of information will better prepare these smaller collections to plan and prioritise their work, especially as regards the challenges of preservation and the actions undertaken by larger archives. In some cases, larger archives may be able to perform certain preservation activities for smaller institutions, including hosting smaller audiovisual file-based collections until digital preservation becomes more widely affordable.

17. MAINTAINING THE KNOWLEDGE BASE OF ARCHIVES

An audiovisual archive relies heavily on the maintenance of an entire system necessary to preserve the documents in its care, and to provide access to their content. Of crucial importance to this system, in addition to specialist equipment and facilities, is specialist expertise and experience. It is a requirement, therefore, that the archive works to equip itself with the necessary skills and knowledge, and to maintain and retain these at a high level. The development and passing-on of expertise to subsequent generations is a particular challenge, requiring careful planning and resourcing.

The archive must, therefore, keep itself and its employees updated with the latest scientific and technical information from the field of audiovisual archiving. This will include information concerning the extraction of both primary and secondary information from carriers, and improvements in preservation and restoration practices.

IASA TECHNICAL COMMITTEE AT THE TIME OF REVIEW:

George Boston
Kevin Bradley
George Blood
Mike Casey
Stefano S. Cavaglieri
Matthew Davies
Carl Fleischhauer
Jean-Marc Fontaine
Jouni Frilander
Ross Garrett (Secretary)
Lars Gaustad (Chair)
Bruce Gordon
Clifford Harkness
Jörg Houpert
Albrecht Häfner
Jean Christophe Kummer
Drago Kunej
Chris Lacinak
Franz Lechleitner
Hermann Lewetz
Xavier Loyant
Brad McCoy
Guy Maréchal
Michel Merten
Stig L. Molneryd
Kate Murray
Marie O'Connel
Bronwyn Officer
Will Prentice
Richard Ranft
Dietrich Schüller (Chair Emeritus)
Joav Shdema
Tommy Sjöberg
Gilles St-Laurent
Bill Storm
Adolph Thal
Nadja Wallaszkovits (Vice-chair)