DIGITAL AV MEDIA DAMAGE
PREVENTION AND REPAIR

Data damage and its consequences on usability

Deliverable D2.1

SEVENTH FRAMEWORK PROGRAMME

DAVID identifier: DAVID-D2.1-INA-WP2-DamageAssessment_v1.20.pdf

Deliverable number: D2.1

Author(s) and company: J-H. Chenot (INA), C. Bauer (ORF)

Abstract: David Deliverable D2.1 reports on digital audiovisual data damage and its consequences on usability. It reports, from inquiries targeted at broadcasters and TV archives, on collected evidence of damage, on the consequences of loss events in terms of the impact on visual and audible properties of the content. Main potential sources of damage are identified depending on their origin analogue, digital, and system-originated problems. Commonly used mitigation procedures are presented. It is stated in the conclusion that complex media asset management systems appear to be more and more vulnerable to risks due to hidden dependencies and interoperability problems, often revealed on the occasion of a migration or a component update.

Internal reviewers: Jörg Houpert (Cube-Tec)

Work package / task: WP2

Document status: Release

Confidentiality: Public

DOI: 10.7800/304dav2

Version	Date	Reason of change
0.01	2013-03-25	Document created
0.02	2013-05-31	Outline
0.03	2013-06-31	Redaction started
0.04	2013-07-12	Complete Draft
0.50	2013-07-18	Advanced Draft for internal review
0.70	2013-07-24	Edited after internal review
1.00	2013-07-30	Released
1.10	2013-08-08	Minor corrections and adjustments
1.20	2013-10-22	Supplement with DOI for publication

Acknowledgement: The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 600827.

1 Tables

1.1 Table of Contents

1.2 List of Tables

2 Introduction

2.1 Purpose of this Document

Digital audiovisual content is everywhere: film, television and online media; personal content from cameras and phones; and content from environmental monitoring, corporate training, surveillance and call recording to name but a few. Preserving audiovisual content in these applications target content reuse, regulatory compliance, and archive monetisation – each with differing needs for content quality, safety, storage, access and budget. But there is a common challenge: how to keep audiovisual content usable in the face of adversity: obsolescence, media degradation, and failures in the very people, processes and systems designed to keep this content safe.

The aims of the European Commission funded FP7/2007-2013/600827 DAVID (*Digital AV Media Damage Prevention and Repair)* project are to analyze the origin of potential damage and its consequences on the usability of audiovisual media content, to monitor and restore damage already happened and will develop strategies for avoiding future damage in a way that balances long-term costs, risks of loss, and content quality. Testing, evaluation and demonstration by archive and industrial partners using real-world data ensures high quality project results. As media we target digital video tape, which for many remains the de-facto form of born-digital audiovisual content, as well as the rapid adoption of audiovisual content stored as files in IT based systems.

Objectives for Work Package 2 *Understanding loss and its consequences* are to clearly define the position of the problems to be addressed by the project, by collecting evidence, documentation, test material, and user requirements, related to digital audiovisual contents, in the view of fine-tuning the scope of the DAVID project. This is done through three tasks:

- Task 2.1 addresses damage assessment data and infrastructures, assessing consequences of loss on usability
- Task 2.2 addresses understanding loss modes, i.e. how loss of/damage to audiovisual content occurs
- Task 2.3 focuses on providing guidance and specifications for the developments and usability aspects of the developments.

Task 2.1 addresses damage assessment data and infrastructures, assessing consequences of loss on usability. The core outcome of this task is the current document: Deliverable D2.1 *Data damage and its consequences on usability*.

2.2 Status of this Document

The current status of this document is Release – Public

2.3 Executive Summary

This document is the Deliverable D2.1 of the DAVID Project prepared under the Seventh Framework Programme (Collaborative Project ICT-2011.4.3 Grant Agreement 600827).

The current document reports on Data damage and its consequences on usability.

It reports on collected evidence of damage, on the consequences of loss events in terms of the impact on visual and audible properties of the content. This report is complemented for internal project uses by a collection of supporting material (video/audio files).

The rest of this report is divided in the following sections:

In Section 3, we document the approach used for obtaining information related to data damage in the audiovisual domain.

Section 4 is dedicated to giving a number or necessary definitions.

Section 5 documents problems affecting digital contents, when caused by their analogue origin.

Section 6 documents problems affecting born-digital contents.

Section 7 addresses System-originated problems, and the procedures for addressing these.

Section 8 is the conclusion.

3 Approach used

The current section reports on the approach used for documenting the cases of damage to video content data, and on their consequences on usability of the contents.

3.1 Studied Archive formats

David is a STREP, a *Specific Targeted Research Project*. Given the limited size and scope of the project, it does not address all the possibilities in terms of data damage, but focuses on the specific domain of audiovisual contents. More specifically, the main focus is on digitised video and film content. The pure audio domain has its own specificities that deserve specific attention, but the boundary had to be set somewhere, it was decided that we would only address audio as the soundtrack for video contents. On the other side, the film industry has its own developments going on, and it was considered preferable that the core focus of the project should be the broadcast production and archives media. A list of media that would be principally considered was established, and is given below.

Format	Comment
Digital Betacam video tape	Ubiquitous
50Mbps MPEG-2 IMX (D10) on video tape	For SD, used in ORF, from production, 600,000 hours backlog to be ingested. Use MXF when exported to files.
P2 cartridges	Contain MXF OP-atom, can store DV, DVCPRO, DVCPRO25, DVCPRO50, DVCPRO-HD, AVC-Intra
MXF Op1a lossless JPEG2000 (for SD)	For SD, ~100Mbps, used in INA
MXF Op1a 200Mpbs JPEG2000 (for HD)	For HD, used in INA
MPEG-2 long GOP PS (8Mbps)	For SD, used in INA
50Mbps XDCAM HD 4:2:2 (RDD9 V2)	Contain MXF OP1a, MPEG-2 long GOP, used in INA, ORF
Matroska FFV1	Used by Archivematica, Austrian Mediathek...
MXF	A large number of different formats make use of MXF.
Apple Prores HD	DCT based, used in production (INA, ORF...)
Avid DNxHD	Typically stored in an MXF container. Used in production (INA, ...)

Table 1 : Non-exclusive list of considered formats

This list is by no means extensive, but should give a good idea of the great variety of the audiovisual contents that could be affected by damage. We will see below that we cannot be exhaustive, and that the same kinds of problems can arise with very different kinds of media.

A trend is visible, that a large fraction of the listed formats use MXF, either at the main wrapper, or as an export wrapper when exporting from tape to files. This already gives an idea of the large number of formats that use MXF... in a great variety of flavours!

3.2 Questionnaires

The first action within the D2.1 Task was to prepare a questionnaire that would serve as a tool for running interviews with knowledgeable people involved in producing, archiving, and exploiting video contents. The questionnaire was intended to be filled, together by interviewer and interviewee. The core of the Q2.1 Questionnaire document was a simple table, with the following fields:

- Problem name
- Frequency (1 - very scarce, to 5 - very frequent)
- Impact (1 - "low impact" to 5 - "high impact")
- Detailed description of problem
- Impact: comments on the consequences/impact (both technical and for re-usability)
- Contact for getting files or tapes samples

The contributors were invited to provide problems described by the fields above, in a number of domains. The suggested domains were:

- Digital video tapes problems
- SDI-originated problems
- Intrinsic quality problems
- Digital files problems
- Human-originated problems
- System-level problems
- Other problems worth mentioning?

Overall 24 people, from 11 different entities were interviewed, principally by INA and ORF. It became soon evident that, while the questionnaire was useful in helping tracking down the damage that could occur to video content, the scope was too wide, and the feedback too diverse, to be able to provide quantitative information (frequency, impact...). Information rather came as series of success (and often failure) stories. Overall, the feedback was however sufficient to help defining ranks in terms of criticality. For example, Bit Rot, although it was often acknowledged that it did happen, was considered as much less critical than System-level problems.

3.3 Collection of samples

At the same time as the interviews were run, a request for samples of damaged content as issued. The request was covering samples of:

- Faulty digital video tapes
- Recorded contents from faulty tapes
- Corrupted Video files

Those samples were collected and transmitted (in most cases as digitised files) to the project partners for evaluation. INA also produced synthetic generated DigiBeta errors for structured problem analysis.

In some rare cases, parts of the contents may become available for wider use, but most of the contents are under Intellectual Property Rights, the samples are therefore not readily available to the public.

The main problems that were collected were digital drop-outs and transmission errors, but we also collected a large number of samples of analogue defects, as we discovered that (at least in the archive domain), a large number of analogue-originated problems are still occurring and limiting the usability of contents. It was also understood on this occasion that a number of artefacts were poorly documented, as mentioned in Section 5, and that knowledge on the origin of an artefact was very unevenly distributed.

4 Definitions

First of all, we must consider that the word "damage" can cover a broad range of different meanings, from physical damage to a cassette, or a very minor drop-out that is not even visible, to the full loss of thousands of files.

In terms of consequences on usability, the same damage can result in a wide range of possible consequences, depending on the affected contents. When a backup copy exists of one (several) damaged programme(s), the main consequences will be additional costs and delays on getting the hold of the backup copy (copies), and restarting operations. Even without a backup, a few frames damaged or lost can have no consequence at all ("we will never use that part"), little consequences ("we can use a copy of slightly poorer quality") or can result in huge losses ("this was the key sequence, it cannot be recovered or re-shot, we cannot use the programme at all"). The consequences will also vary depending on the user: a broadcaster will have different views from a producer, a news editor, an archivist, or a footage library.

We have tried in this section to set up a number of definitions that will hopefully help clarifying the scope of our topic, keeping in mind that we have tried to get as general a feedback as possible from the users, without insisting on one kind of damage, or on one kind of consequences to the users.

The table below includes a number of different issues that can have a consequence on usability, some of which can be considered as 'damage'.

Note: this table is freely inspired from the list of issues defined by the Open Planets Foundation, in their Digital Preservation and Data Curation Requirements and Solutions page[1], and was adapted for our needs, and extended to include the physical origin of the problems.

Issue	Description
Missing carrier	The physical artefact that holds the AV content has disappeared
Damaged carrier	The physical artefact that holds the AV content is damaged
Missing playback machine	There is no machine for reading the carrier
Incomplete playback	A section of the programme is missing, possibly to damaged carrier
Poor playback quality	The quality of the playback of AV contents is insufficient
Encoding quality issues	AV asset is not of required quality, e.g. due to low bit rate compression
Bit rot	AV digital data has become corrupted
Block read error	A section of the digital medium was not read properly, resulting in distorted parts of the AV contents
Conformance issues	AV data does not match the required profile or the file format is unknown
Context problem	Required environment/metadata to understand the AV data is missing
Embedded objects issues	Interpretation of other objects embedded in digital AV file is difficult
External dependencies	Exploitability depends on relationship with objects outside the digital AV file

[1] http://wiki.opf-labs.org/display/REQ/Digital+Preservation+and+Data+Curation+Requirements+and+Solutions

Structural relationship issues	The AV asset is made up of more than one digital object, and structure is unclear
Rights issues	Permissions for using or migrating AV asset are unclear/missing
Obsolescence issues	Format, software, hardware is no longer available/understood

Table 2 : Issues and Damage

From the problems above it is possible to set up a list of different failure modes, detailed in the table below:

Affected system part	Failure mode
Equipment/hardware (e.g. tape drive, server, cables)	Equipment failure (HDD head crash, tape head clog), server failure, network failure / chattering connection (e.g. for NAS)
Carrier (e.g. tape, HDD, optical media)	Whole carrier corruption, block-/sector-level corruption, burst, bit flip, misdirected write
File system (e.g. NTFS, ZFS)	Corrupt file system records
Operating system	I/O failure, corruption of data in memory
File: format, container, codec	Inconsistent/non-standard format; truncated; corrupted
Application	Interprets standard poorly/differently to expected

Table 3 : Failure Modes

From the problems above it is possible to set up a list of different types of loss, from the most harmless, to the complete impossibility of using a large set of contents. The following table gives a first list of the different failure modes, ranked from the lowest to the highest severity.

Type of loss	Examples
Can play now, but may not be able to play or migrate in a different context	Metadata is ignored on playback but used during transcoding (such as width and height, which is ignored on playback but causes conversion to change aspect ratio) Dark or reserved metadata (that may be corrupt or wrong and may be ignored by current player) is interpreted differently by another player (now or in the future)
Can play, but there are artefacts	Non perceivable artefacts Hardly perceivable artefacts Frame dropouts Corruption of frame(s) Corruption of audio stream (whole/part) Missing parts (e.g. subtitles, 2nd audio track) Wrong colour space Wrong aspect ratio
Can play, but information about the asset is missing/wrong/corrupt	Metadata is missing or incorrect, meaning that we cannot ‘access’ the asset easily (e.g. using search)
Can play (technically), but do not have rights appropriate to intended use	No rights to broadcast, provide access to, or migrate content
Cannot play without additional processing (we know how to do it)	Must first transcode from obsolete or archive format
Cannot play, but a backup copy is available	Must use the backup copy, may involve important efforts

Cannot play without additional processing (it should play)	File is incompatible with current stack, despite the fact that it should be in a playable format File was created ‘bad’ (internally inconsistent), e.g. not according to the profile it says it is (OP-Atom, not OP1a, 8bit not 12bit) We upgraded the stack but did not change the format/carrier, on the understanding that we should still be able to read it
Cannot play, format is obsolete with respect to current stack	We waited too long to migrate the content. Migrating content now is difficult because context required to interpret the file is obsolete/prohibitively expensive to replace
Cannot read the file	Equipment and/or software required to read the file is obsolete Carrier is damaged (head crash, tape problem) File system (e.g. journal) is corrupted Other problems resulting from previous migration to different file system/carrier – e.g. non-standard character used in filename

Table 4 : Types of loss

5 Analogue-Born content problems

When the inquiries were started, it was preferred not to orient the answers towards the problems affecting digital-born contents, but rather to allow the interviewees to express all the problems that affected the usability. It is stressed here that a large proportion of the users declared that, despite the contents they were using were digitised, they were still plagued by problems due to the analogue origin of the contents. It is not surprising that those that were the most insisting on these topics were the personnel involved with managing of using archive AV contents. Still, the availability of digitised analogue-originated contents makes even more visible the lack of adequate tools for managing some of the problems that affect these contents.

We will not try here to set up an exhaustive list of the problems that can affect analogue-originated archive contents. Such lists have already been set up, we prefer to direct the readers to several sources of information:

The Aurora/Brava impairments dictionary[2] ; despite being quite old (earliest version is dated 1996), it offers a large (bi-lingual) palette of the different origins and defects that can affect film, audio, and video contents. Missing in this list are pictorial samples or short videos to demonstrate the effect of each of those artefacts.

Although currently less detailed, the AV Artifacts Atlas (AVAA)[3], set up in a joint effort by the Bay Area Video Coalition, New York University Digital Library Technology Services, and Stanford Media Preservation Lab, aims at defining collaboratively an on-line common vocabulary on analogue- and digital-originated video and audio artefacts. AVAA has much more effort currently running, and uses modern technologies (MediaWiki), and has pictures for helping understanding the different artefacts.

We also have to mention the EBU effort by Strategic Programme on Quality Control (QC)[4], that is currently working on a set of criteria of acceptance of broadcast contents. The list is under review at the moment, but does not list artefacts as such, but rather the criteria that have to be measured.

Fortunately there are many analogue-born content problems that can currently be addressed by state-of-the art techniques. In the case of David, we will restrict the Analogue-Born content problems to those that are involved in restoring contents have expressed as being currently unsolved. Problems listed as still pending are listed in the table below (in alphabetical order). Descriptions are adapted from the he Aurora/Brava impairments dictionary.

Name	Description/comments
Antiphases	This artefact, only present in SECAM, results in a permutation of DR and DB components (from a few lines to a whole field are affected). May be due to loss or attenuation of field identification signals, loss, attenuation, or incorrect timing of colour burst. The colour is considerably shifted towards pink.
Blur	Blur is very common, and may be caused by several factors (incorrect setup of tube cameras, poor lens focus, motion...). The tools currently available for video restoration are not efficient.
Beta and BetaSP chrominance head clogging	Dirt on one chrominance head: only one field has chrominance signal. There is no tool for repairing this artefact, despite it is relatively frequent, as chrominance heads are more exposed to dirt than luminance head, and the problem was not visible on B/W viewfinders and monitors.
Comet/blooming	Memory effect on tube cameras, affecting areas exposed to too much light. No tool is currently available for correcting these artefacts.

[2] http://brava.ina.fr/brava_public_impairments_list.en.html
[3] http://preservation.bavc.org/artifactatlas/index.php/Table_of_Contents
[4] http://tech.ebu.ch/qualitycontrol

Noise/Grain	Video noise and film grain are very frequent in archive documents shot on video and film. There are a number of tools available, but the quality of the result is not as good as the expectations.
Overshoot, Ringing, Echoes, streaking	A number of different defects originated in electronic circuits and transmission cables, that makes sharp edges look oscillating, softer, or replicated, in the horizontal direction. No tool is currently available for correcting these artefacts.
Several pictures unrecoverable.	Several Frames damaged beyond repair, due to mechanical tape Problems, damaged film, film breaks, off-locks... Despite there are tools for repairing one picture, when several pictures in a row cannot be recovered, there is usually no solution available.
Smearing	Saturation of CCD cameras resulting in charges flowing between adjacent pixels or registers. The consequence is a saturation of one colour channel extending along rows (Red channel is usually affected first). No tool is currently available for correcting this artefact.

Table 5 : Analogue-originated problems

The good point on the analogue-originated artefacts is that, as analogue production has nearly fully stopped, the artefacts list is not supposed to extend dramatically beyond the known problems. Digitisation efforts have started by many contents holders, and the process, albeit costly, is expected not to change significantly over next years. We do know how to address most problems, for the others we know that we don't. However, this should not be considered as a reason for slowing down the digitisation efforts, as the quality of the playback is not expected to improve very much in the next years, at least it will not improve fast enough to compensate for the decaying rate of the analogue contents.

6 Born-Digital content problems

The introduction of digital video, starting from 1982, has created a number of problems different from those mentioned above. We will in this section address first the case of digital video tapes, then other digital media, then transmission problems, and eventually consider the case of files.

6.1 Digital video tapes problems

The history of digital video tapes is briefly summarised here.

All the video tapes systems mentioned below use helical recording technology. In Standard-Definition (SD), the first digital video tape format to be introduced, in 1987, was D1; it stored 4:2:2 video contents in Standard Definition (SD) at 8 bits per sample. The very high price, and interoperability problems between brands, made the number of units sold relatively small. D2 (1988) and D3 (1991) stored 8 bits digitised composite video, and were not widely adopted, with some exceptions (e.g. BBC). Digital Betacam was introduced by Sony in 1994. It was DCT-based, used a relatively low compression rate (2.2:1), and could store 10 bits 4:2:2 video with a quality very close to uncompressed, with robust VTRs and relatively affordable units that made Digital Betacam was quite quickly adopted by the broadcast industry as a *de facto* standard, both for cameras and studio recorders/players. D5 was a good 4:2:2 10 bit uncompressed format from Panasonic, launched in 1994, but was by far not as successful as Digital Betacam. A high number of different formats emerged then, all based on DCT compression: Sony built upon their success with the Betacam SX (1996, 8 bits 10:1 MPEG2-like IBIB compression), that was quite successful in the news gathering domain. In 2001, the Betacam IMX format, with pure intra-frame encoding compression rates as low as 3.3:1, was the last Sony SD video tape recording system, and became quite widespread. In addition most IMX decks could play Digital Betacam cassettes, and could transfer IMX tapes faster than real-time. The consumer format DV, introduced in 1996, evolved towards relatively more robust formats: DVCPRO (Panasonic, 1995), DVCAM (Sony, 1996), DVCPRO50 (50 Mbps, 1997). All DV-originated use DCT-based compression, had small cassette sizes, low tape thickness, and were mainly used for acquisition, their long-term performances being often criticised; but the contents could be directly transferred into files. Digital-S was an alternative offered by JVC, with limited success.

For High-Definition, the Sony HDV-1000 (1-Inch tape format, 1984) was not digital, but analogue. Early digital systems using 2 to 4 D1 VTRs were used, but the reliability was quite questionable, and recovering the contents from the originals has become extremely difficult.

D5-HD, a High-Definition upgrade for D5 was proposed in 1994 by Panasonic, using compression, followed in 1997 by Sony HDCAM (8-bit DCT compressed 3:1:1, 144 Mbps), and in 2003, of HDCAM-SR (10bits, 440/880 Mbps, MPEG4 part 2). Again, Sony has been quite widely distributed, despite the high price of HDCAM-SR. Alternatives were offered by JVC (D-9 HD, based on Digital-S), but again with limited success.

The digital video tapes exhibit problems similar to those of the analogue world, with different manifestations. A **drop-out** is a loss of Radio Frequency (RF) signal read by one (or more) of the microscopic magnetic playback heads, reading at high speed (several m/s) the recorded magnetic track. The drop-out can be caused by one or several of the following causes:

- The tape was briefly lifted away from the head by some debris, or by a tape crease.
- The head is not properly aligned with the magnetic track
- The magnetic track is damaged, e.g. due to magnetic debris on the tape path (magnetic scratch), or age.
- The RF signal on the tape is too low, due to a drop-out at record time, poor settings, damaged recording heads...

In the case of digital video tape, the picture contents are often compressed, and transformed into data blocks that are written on tape as parallel diagonal magnetic tracks. A very small drop-out will result in one or several blocks read in error. A large drop-out will result in a very large number of blocks not read at all, and visibility will be worsened by the level of compression. To limit the visibility of those accidents, the manufacturers have devised and implemented a number of strategies:

Error correction: if a few bits are in error, Error-Correcting Code (ECC) disseminated in same or other blocks can allow re-computing fully the block. Correction should be perfect, loss is very unlikely.

Error detection: a checksum or cyclic redundancy checks (CRC) disseminated in same or other blocks can allow detecting the block is in error, error concealment will be activated.

Error concealment: if ECC are not sufficient, the video tape playback system will attempt to conceal the problem by replacing the affected pixels by values computed in an attempt to hide the problem, in general using pixel data from adjacent areas in the same field, or from same location in the previous field(s). This strategy is widely used by digital tape playback systems; and is often quite imperceptible, provided that the drop-out is short enough, and isolated. There are different kinds of error concealment; recent playback decks often have an improved behaviour than older ones.

Shuffling: to avoid propagating errors if concealment always takes place at the same place, e.g. in the case of tape crease, or magnetic scratch, several systems (Digital Betacam, D5, HDCAM...) have installed a pseudo-random order in the way the blocks coding for different areas of the picture are organised on each track. This has several benefits:

- Error concealment will hopefully use clean data for replacement
- Affected areas are more numerous, but smaller, making error concealment, or restoration, easier.

The DV (DVCAM and DVCPRO) families don't use shuffling, this makes playback worse when too many drop-outs arise, as affected blocks are organised as contiguous horizontal lines.

Head-clogging occurs when debris accumulates on the head, tape-to-head distance becomes too important, and the drop-out lasts until debris are evacuated... or a cleaning procedure is applied. When head-clogging arises, continuing playback is usually useless.

A **mitigation procedure for avoiding drop-outs and head-clogging** when playing back cassettes is to **clean the tape** using a suitable cleaner device; cleaning the tape path and the head on the player is also necessary. When the drop-outs are not recorded on the tape, it is possible that reading the same tape a second time may not cause drop-outs at the same position on the tape, thus recording several playback attempts from the same tape can contain enough good content for re-editing the full programme from the different fragments. Of course, if another good copy is still available, the faster option of using the alternate copy may be preferred.

It may be worth to be noted that tape baking (heating the cassette at 50-60 °C and low humidity for several hours in a dedicated oven) is considered as inefficient and dangerous for digital video tapes contents, as the tape formulation uses metal particles (MP), or metal evaporated (ME) technologies). Tape baking is only useful for analogue oxide tapes (e.g. U-Matic, Betacam...).

6.2 Other digital carriers problems

There are other digital carriers than digital video tapes. One could argue that every media programme has to reside somewhere on a RAID, on a server, or in cloud storage. These will be considered in Section 7. Here we will address digital carriers where every item can be clearly identified as residing on a physical medium.

Dedicated media such as P2 cartridges (flash memory) and Sony Professional Discs (Blu-Ray-based) are used quite widely in the news gathering and production activities. Both systems use MXF as a wrapper, organised in a tree structure of files, holding metadata and audio (P2). Accessing the contents usually involves opening the contents as any removable storage media, and importing the contents. Both systems are recognised as quite robust, but, as video tapes, they share a common intrinsic weakness: between the time footage is shot and the time where the contents are copied onto an editing system, they contain the only and sole copy of the footage. The principal risk associated with this media lies in being damaged, lost, or stolen within this time interval, before being copied. We have also received complains about file and system incompatibilities problems; these are reported in Section 6.4.

Beyond dedicated digital media, **any kind of removable device** (memory card, USB key, external USB drive, DVD-ROM, data cartridge) can be used for storing video files. When this is used for primary acquisition, the same risk (being damaged, lost, or stolen before being copied) is present. When the medium is used for transporting contents from one point to another, the source is often still available, and the lost is only expressed in time and cost for coming back to the source.

Data tape cartridges (e.g. S-DLT, LTO1 to LTO-6) can be used as dedicated media, on shelves, or within a robotic library. Files are usually not distributed on several tapes. Risks associated with such media lie in destroying of damaging a cartridge (dropping a cartridge on a hard floor may result in total loss of the contents, a tape may be eaten by a drive), interoperability problems (the expected playback compatibility between drives of different generations seems to be only partially verified, as an increase in the number or irrecoverable errors was experienced).

6.3 Digital video transmission problems

Some cases were reported to us of rare cases of recorded contents affected by transmission errors on SDI (Serial Digital Interface) cable links. Such errors were caused by cables longer than the standard acceptable length (300 meters), or damaged. The degradation in the picture should be quite brutal, from a few pixels in error, to a complete disappearance of the contents, but we have found cases where only some of the SDI words would be distorted, resulting in pictures that would be correct except at a specific luminance level.

Cases were also reported where people walking in a technical room would induce brief signal interruptions in BNC connectors. Such problems usually result in series of pixels affected, within the same few lines.

6.4 Digital video files problems

The era of digital video tapes is coming to an end, with a dramatic reduction of the number of hours produced on digital video tapes, and the corresponding increase of contents stored on optical media, or directly on Hard Disk Drives (HDD). With the exception of the early stages of production, where the contents are stored in cartridges (e.g. Panasonic P2, or Sony XDCAM disks...), the contents are usually replicated as files onto a server before being edited. The result of the editing becomes a new file (or set of files), and it is less and less frequent that the edited contents at the highest quality would be reported onto a digital video tape. A consequence of this is that there are many contents that solely exist as one or several files distributed over a number of systems, at a much lower cost, but causing concerns on the reliability, and on the possibility of accessing the same file after several years, or even decades. This section addresses the problems that can affect those files.

Bit Rot is a general term for designating random corruptions or one or several bits in a data storage device. The corruption could take place in the media themselves, in the electronic storage and transmission circuits, in computer memory while the contents are used, in solid-state storage... The causes can be excessive light (optical media), various decay processes (solid-state storage), cosmic rays... The frequency of such corruptions should be extremely rare, but there are concerns, with miniaturisation, use of non error-corrected subsystems (e.g. computer memories), the ever-increasing amounts of stored data, and the increase in file size (a 1h-hour programme of HD 4.2.2, encoded in JPEG2000 at 200Mbps represents 90 Tbytes of data), that even unlikely single bit flips could become in future years a major cause for file corruptions. When running the interviews, we have been open to claims that bit rot could be a problem, and have also asked for a position on this question. In most cases, a clear answer was given, that bit rot was not a problem at the current time. It was acknowledged that some rare cases or corruptions might have been explained by the occurrence of bit rot, but the importance and the risk of this phenomenon was at the present time much lower than any other possible causes of content losses.

File crash will be used to describe cases when a file is corrupted, or cannot be read. Reasons may be hard drive or data tape local failure, accidental partial (or total) erasure, human error... In those cases, when the file is retrieved, a section of the file is corrupted, the file is truncated, or the file cannot be read at all, since a subsystem will not allow a partial delivery. The interviews tend to confirm that such file crashes are currently much more frequent than bit rot cases. The principally used mitigation procedure here consists in falling back to the backup copy(ies) of the affected file(s).

Full file loss: a full file loss takes place when a file crash renders the file completely useless, because it cannot be read, or an essential section of it cannot be recovered, and there is no backup copy. Isolated full file losses without a backup were repeatedly reported as very scarce. Several interviewees told us "we have never lost a single file".

Backups everywhere: despite a number of file crashes were reported, very little cases of full losses of files were reported to us. In most cases, the principal reason was that a backup copy of the file existed somewhere, and could be accessed to re-generate the file. In some rarer cases, the file itself was lost forever, but earlier copies, possibly analogue, still existed, and a new file could be re-created to replace the lost file. Our findings showed that, in most cases, rather three than two copies of every file co-existed, one of them easily accessible from a backup site, the other ones usually on data tapes on shelves at a remote location. The main exceptions were the off-camera production media, where only one copy existed until safely transferred to the editing stations. The risk for the content is of course maximal during this short period of time.

Interoperability problems were reported as more frequent than file crashes. Indeed, the number and variety of file formats and wrappers has exploded, and the competition between manufacturers often contributed to difficulties when migrating contents from one system to another, or when trying to use a file. Amongst the most frequent interoperability problems:

- Incompatible implementations by different hardware and software manufacturers of a supposedly common standard and profile. Such incompatibilities can result in software crashes when opening a file, software refusing to open a file, incorrect aspect ratio / field order...
- Changes in the contents, due to incorrect translations between colorimetric spaces, at one point of a long transmission chain.

Interoperability problems are quite often caused by implementation or parameterisation errors: most common cases are:

- Non-consistent Presentation Time Stamps (PTS), with several images for the same PTS, or missing images.
- Wrong/inconsistent/mixed aspect ratio
- Wrong/inconsistent colorimetric spaces
- Wrong/inconsistent field/frame wrapping, field dominance
- Valid, but unusual resolution (e.g. 704x576)

Interoperability problems are often latent, and arise only when a new setup is established, with new software, new tools. In such cases, a large part, or all of a large set of files suddenly become unavailable to the new tools, until a solution of found. Possible mitigation procedures for interoperability problems include:

- Updating old software/hardware
- Adjusting the new software
- Modifying the files to make them compatible
- When possible, working around minor (colour/aspect ratio...) problems when using the contents.

Such mitigations procedures are often costly.

MXF-related interoperability problems represent a large subset of the compatibly problems met. MXF is a very complex and widely disseminated standard, and the number of allowed possibilities is such that not all cases can be tested against a setup. The two main Operational Patterns (OPs) used in acquisition and archive domains are OP-atom (e.g. Panasonic P2, Digital Cinema Initiative...), and OP1a (Sony XDCAM, video archives...). Those Operational patterns are well defined, but, as MXF is a wrapper format, very flexible, few constraints are enforced about the encoding of the contents, every manufacturer proposes his own profiles, and nearly every user has his own profile, or sub-profile. Ingest systems and tools quite often expect only specific profiles, or try to address a very wide range of settings, but cannot test every situation, as these are too numerous. There is usually a limit to the load that a system is able to cope with (file size, bit rate, codec flavours). When this limit is reached, degradation occurs, e.g. playback is not real-time any more, or errors arise. Examples of such problems, obtained during a interoperability check campaign (INA, end-2010) are given below:

- Software crash
- No playback
- Playback stops
- Playback is jerky, or slower than real-time
- Colour flashes (colour coding compatibility problems)
- When OP-atom is used, a programme comes as several files (.MXF for video, one or several .WAV or .MXF for audio), that have to be re-integrated into one programme
- Contradictions between MXF descriptors and the actual encoding, causing a number of malfunctions
- Wrong aspect ratio at playback
- Squashed picture cased by field wrapping options incompatibilities.
- No audio playback (unsupported audio encoding e.g. AES3)
- Flashes due to wrong frame rates (59.94i instead of 50i)

Some of the problems above could be solved later by updates from different manufacturers, but it was found that that many MXF files do contain contradictions, and happen to play by accident!

Indeed, MXF was made from the beginning to accommodate a large number of uses, from recording on-the-fly single-shot files to transmitting complete edited stories, allowing for a large number of (video, image, audio) codecs. This results in a powerful standard where the same information may be written at different places, where there are ambiguities as to the meaning and the importance of a number of parameters, where the complexity of the standard increases the possibility of incorrect implementation by hardware and software manufacturers. In our experience, the IRT MXF Analyzer tool finds errors, problems, and inconsistencies in most, if not all, analysed files. However only a few of these errors have a real effect for the media playability; those files play, or seem to play properly… By contrast, it is possible that an MXF file that appears to be valid, would contain content (audio, video, images) that have an encoding problem, or where the encoding is inconsistent with the MXF parameters (resolution, GOP, lossy/lossless…). To fetch this information may require not only analyzing the wrapper, but also decoding the whole file…

The most currently distributed **file-based recording media** making use of MXF appear to be Sony XDCAM (actually 5 completely different profiles: DVCAM, IMX, XDCAM HD, XDCAM EX and XDCAM HD422, XAVC), and Panasonic P2 (actually 7 completely different profiles: DV, DVCPRO, DVCPRO25, DVCPRO50, DVCPRO-HD, AVC-Intra, AVC-Ultra). But there are many other different profiles, designed by manufacturers, or customised by users for their own needs.

It is worth noting that both XDCAM and P2 have their own way of organising the contents as files in a tree structure, including MXF and other (XML, images, editing instructions) files. Some problems arise when the structure or the contents are slightly different from expected, then resulting in impossibilities to easily recover the full contents of the media. Cases were reported to us where the use of special characters (`"/,@?!), foreign character sets (e.g. Cyrillic) in would crash the ingest system. In other cases, programmes spanning over several media would result in one of the files silently not being ingested.

To address the problems caused by the too wide potentialities of MXF, the Advanced Media Workflow Association (AMWA) has endeavoured setting up a number of Application Specifications, and specific subsets (shims) that would cover specific needs for different uses : are published Versioning AS-02, Program Delivery AS-03, Production AS-10, Program Contribution AS-11, Commercial Delivery AS-12. Archiving & Preservation AS-07 is currently work in progress. It is still unclear whether enforcing compliance with such Application Specifications will help reducing the cases of interoperability problems.

7 System problems and mitigation procedures

7.1 The system risks

Up to here, we have reported on the problems and cases of content data loss caused by the individual subsystems. But we have been told that the worst cases of trouble and data loss in the digital domain were actually not caused by the subsystems themselves, but by the assembly within complex systems of a large number of components, e.g. Media Assets Management (MAM) systems. The concentration of large amounts of data that had to pass through an ever-increasing number of parts, seem to increase the risk in often unexpected ways. The problems reported to us belonged to the following categories:

- Human errors
- Hardware incompatibilities
- Incompatibilities between subsystems
- Failure of central control systems
- Migration problems

Human errors are a major cause of concern. It can be argued that most of the other categories may also be caused by human errors (e.g. poor code, incomplete checking...), but we will concentrate here on direct human errors. In any complex system, operators have to be in charge. They have to perform essential tasks, maintaining the system in operation, checking that resources are sufficient to face unexpected conditions, and recovering the problems that can arise. However vigilant an operator is, he will always make errors, usually without consequence, but sometimes for the worst. The list is virtually endless, but one can cite:

- Removing more files than wanted
- Removing files in the wrong folder
- Pulling out from a RAID a working disk instead of the faulty one
- Copying and editing a configuration file, not changing all the necessary parameters
- Editing a configuration file into a bad one, having no backup
- Corrupting a database
- Dropping a data tape / a hard disk drive
- Introducing an adjustment with unexpected consequences
- Replacing a correct file or setup from a wrong backup.

Such errors have the potential for affecting durably the performances of a system, and are not always reversible. In addition, the risk of error is increased by the stress introduced by urgency, e.g. when trying to make some room on in storage facilities approaching saturation, or introducing further errors when trying to recover using backup copies.

Problems caused by **hardware incompatibilities** were reported to us several times: systems that should have been compatible, actually weren't, or did not exhibit the level of reliability that was expected from compatible systems. E.g. we have found two cases where data tape drives of next generations, supposed to be able to read earlier generation tapes, did read them, but with unexpected effects: in both cases, the level of errors would progressively increase up to unacceptable levels. In one case, reading earlier tapes on next generation drives did cause a quite accelerated head wear on the drive, contributing to a disruption of service over a complete data tape library. Interoperability problems between data tape drives (LTO-3) from different manufacturers, fully compliant with the same standard, were also reported to us. It was also reported to us, on rarer cases, that one faulty data tape drive would, un-noticed, consistently ruin a complete set of tapes, thus making the data unrecoverable.

Incompatibilities between subsystems can also generate a whole range of problems, which sometimes go un-noticed until very late. Interoperability problems between video encoding and decoding subsystems were cited above (headers, metadata and file names with non-ascii character sets, colour spaces, interlaced/progressive, field dominance, aspect ratio, resolution...). Such problems can cause inefficiencies, or worse. When the incompatibilities are detected at setup time, they can delay the entry into service. If not, the processes will start, generating new data, new files, that cannot be used immediately, as they are incompatible with the tools that should use them. This can have pernicious effects, as the users, under the pressure of urgency, may try to find ways around the problems (e.g. correcting a wrong aspect ratio with tools at hand, or locally transcoding the file), not always reporting about these, thus allowing for work being done with more effort than should have been necessary; but the flaws may go un-addressed before years after system entry into use, e.g. when a new component has to be added to the system.

When an action is decided to fully correct the problem, considerable efforts are already wasted, a large number of files may have to be patched, and a number of dependencies on these files may result in unexpected behaviour when trying to re-use the same file.

Beyond those incompatibilities, Media Asset Management systems rely on an increasingly complex set of resources, and on critical functions, and can fail dramatically if some conditions are not met. **Failure of central control systems** may happen, e.g. the failure of one of the following services can make whole systems fail:

- DNS (Domain Name Server)
- User authentication system
- Email service
- Database
- Software licences on dongles
- Power or disk failure on any server
- ...

Disruption of any of these services should only prevent the system from working until the service is restored. But such dependencies are not always fully documented, and it can be difficult to track down the origin of the problem before restoring the service. Even there, restoring the complete system to normal operations can be lengthy, and it is not 100% sure that nothing will be lost.

The risks are often amplified (or revealed) by **Migrations**: from one system to another, from one data tape type to another, from one storage system to another, from one encoding standard/profile to another. Indeed, migrating a set of files or a system to other facilities relies on the capacity of reading the totality of the source data, and even more, on the capacity of using these data. But an existing **-** up and running **-** system can run for years with minor problems, this does not give a guarantee that all data are available for copying. Latent errors are probably there, i.e. errors that will only become visible when accessing the files. Such errors may be in limited numbers if the data reside on well-maintained RAID arrays, but are more likely to occur if the data reside on data tapes in a library. On a system where a small part of the contents are accessed every year, copying all the data in a few weeks or months will increase the number or errors by increasing the amount of read data.

Migration onto a new system can also involve **migrating the files into a new format**. There are a number of incentives on doing this: achieving a better compression ratio with same quality, getting rid of obsolete formats, ensuring a better interoperability, faster access... However these advantages have to be balanced with problems that may arise: in most cases the new files will not be readable by legacy systems; the old files will have to be maintained as well, or removed, sometimes with unexpected consequences. Migrating to a new codec without significant loss of quality may actually increase the required file sizes... Overall, it was reported to us that it was considered, at least for master files, **safer to stay with the same media files** as long as possible, i.e. as long as codecs are available, only generating new files with superior performances (resolution, quality...) for newly ingested contents, and having systems managing both old and new files profiles as long as possible.

Migrations can also reveal interoperability problems: adding a new subsystem, changing or updating an existing subsystem, or moving to a new system can reveal unknown or new incompatibilities and dependencies that can block or slow down processes, or make a large fraction of the data and files unavailable. We were given numerous examples of such cases:

- Changing the data tape drives to next generation drives, resulting in increase of errors
- Removing old browsing files to save storage space, and discovering - too late - a hidden dependency on these
- Upgrading some software : essential working plugins are not compatible any more
- Delivering video files to a new customer : despite valid, the ingest system cannot read them
- On the new system, aspect ratios / frame rates / other are wrong
- Some files cannot be read on the new system

Migrations are however sometimes necessary, systems have to evolve to be able to interact with a changing environment; therefore not migrating may not be an option. But keeping in mind that any change or move or update in a complex system, will generate problems, may be an incentive to make such moves, as much as possible, progressive, documented, and reversible, to allow for solving the problems before they become untreatable.

7.2 Mitigation procedures

Having listed above how things could possibly go wrong at the system level, we must also mention how such mishaps can be (and are, in most cases) avoided. We have identified six widely used strategies and mitigation procedures:

- Backups everywhere
- Checking against poisoning
- Repairing
- Ensuring reversibility
- Fallback and recovery mechanisms
- Disaster recovery Plans

Backups everywhere: it is much easier to repair or reconstruct part of a system when recent backup copies of the essential parts exist. This includes not only the media files, but also the source codes, the scripts, the configuration files for every client and server, the databases, the help tools... In terms of hardware, backup copies are also necessary, this includes spare drives, spare RAIDs, spare disks, spare servers ready for fall-over, instructions for re-constructing servers, change-over procedures... In the specific case of audiovisual contents, our findings have shown that for every media file, in most cases **three copies** were made: one - on-line or near-on line - for use, and two as a backup (sometimes the third copy simply resides on disks or tapes from a previous subsystem – before migration). Depending on the configurations, the actual level of redundancy may be higher than three, as one or several of these copies are stored on RAIDs. With this level of redundancy, all the actors interviewed have confirmed they had never lost a single irreplaceable file. The complexity of the backup and recovery procedures may however sometimes be daunting.

Possibly due to a bias in the addressed user base, we have found little evidence in the broadcast domain of the use of cloud storage for storing audiovisual contents at high quality, probably because the actors in the domain consider that networking costs and cloud-storage costs would be still too high. The emergence of such uses may introduce changes, but it is likely that the first adopters of redundant cloud-storage for broadcast contents would still want to keep one or two backup copies on data tapes at a remote location.

Checking against poisoning: Files with errors or wrong parameters can work perfectly well within an existing system, but can make the use of those files much more difficult when exporting to a new system, when updating subsystems, or when adding parts. Therefore it may be well advised to try anticipating for trouble before it arises. Running quality checks at ingest time is of course useful, but after some time, it becomes difficult to track all the possible ways that could have generated a file. Therefore, trying to run quality check, using recent tools, on oldest files, may lead to discoveries that may be useful. More specifically, every running archive system that we have known has samples of files that have internal discrepancies and inconsistencies, or that display the wrong parameters. Running such checks, and collecting user feedback on wrong parameters, is the first step towards correcting the problems, and avoiding later issues.

Repairing: When evidence is collected on the problems that affect the files, it is possible to consider correcting the files. This is not always straightforward, as it involves replacing each file by a patched or re-encoded version, which is not always possible. Generating a corrected version of the file can be considered, provided that it is verified beforehand that not too many tools depend on the file being... wrong! Making a better version of each file in a store could be costly and risky, since it may involve destroying each earlier file, bringing risks if it is discovered later that each patched version is itself flawed! It may be a safer option to document the errors, use workarounds, and take the occasion of a migration to patch the files on the same occasion.

Ensuring reversibility: whatever action is started to mitigate the risks or to correct an error, it is preferable to take precautions to allow for coming back to a previous stable state. This involves verifying that the backups mentioned above exist, and are at hand, preparing snapshots of subsystems before attempting any upgrade, upgrading hardware and software only progressively, testing thoroughly, letting enough time for new setups to prove their reliability before committing for the other upgrades; after a migration, keeping the previous systems and subsystems up and running, or at least ready for re-use, with available and up-to-date documentation...

Fallback and recovery mechanisms: to maintain an online archive up and running with a minimum level of reliability, redundancy is necessary at many levels. Media files have to be duplicated, so that the backup copy can be accessed if necessary. Every subsystem has to be available at least in duplicate. For most critical elements, depending on the necessities of service, instant fall-over mechanisms may be necessary. Such fall-over mechanisms have to be tested and validated before becoming really necessary. As for the parts that cannot be duplicated (e.g. data tape libraries, main network, power lines), those have to be validated as extremely reliable, possibly working in degraded mode while being maintained, or supporting a short interruption of service (e.g. using Uninterrupted Power Supply to resist brief power failure, or being able to replace a key network switch in a few minutes).

Disaster Recovery Plans: in the domain of audiovisual contents, performances are often linked to the existence of a number of integrated systems on one site. This introduces a weakness that will only become apparent when a disaster takes place. A Disaster Recovery Plan will prepare and document the procedures to re-start activity, in case one of the sites becomes partially or totally unavailable due to a disaster. In addition to backup copies of the data being readily available on a distant site, such procedures also rely on the possibility of re-starting the service using other facilities, within a pre-defined *recovery time objective*. A full Disaster Recovery Plan is very complex and costly to set up.

8 Conclusions

We have explored within this document a number of ways in which preserving digitally, and accessing audiovisual media documents, could go wrong. We have first requested a number of stakeholders in the audiovisual broadcast domains, about their experiences in damage to data, and their consequences about usability.

We have used questionnaires, but have found that experiences were quite diverse, and covered a very wide scope, from archive restoration operators that would still be struggling with some artefacts caused by the analogue origin of the contents, to risks of losses due to the failure of key components in the Digital Media Asset management systems.

We have tried to address those problems in three chapters - analogue-originated contents, digital-born contents, and system problems.

In the course of our study, we have identified problems with incidence and consequences lower than expected.

- Although Bit Rot did happen, it was by far one of the less important problems they had to face in the present period.
- Although files could be lost, it was nearly always possible to re-generate the file, from backup copies, or from intermediate copies.
- Cloud storage is still not used very much, at least in the broadcast domain.

On the other hand, problems that generate stress, delays, and possible losses are the following:

- Interoperability problems start to plague the processes at many levels, and are often discovered late enough that consequences become major cost factors.
- Complex media asset management systems appear to be more and more vulnerable to risks due to hidden dependencies, often revealed on the occasion of a migration or a component update.

It is possible that such findings only would have their validity within the audiovisual contents domain. We still hope that those conclusions will help preparing safer and better environments for the long-term preservation and use of audiovisual contents, and hopefully beyond.

9 Glossary

4:1:1	A colour sampling pattern, where chrominance is sub-sampled horizontally: only one pixel out of four holds the colour information. Used by DV in ~60Hz environments (US, Japan...)
4:2:0	A colour sampling pattern, where chrominance is sub-sampled horizontally and vertically: only one pixel out of two holds the colour information, only one line out of two. Used by DV in 50Hz environments (Europe).
4:2:2	A colour sampling pattern, where chrominance is sub-sampled horizontally: only one pixel out of two holds the colour information. Used by most SD and HD recording formats.
4:4:4	A colour sampling pattern, where colour is not sub-sampled horizontally: every pixel holds the luminance and chrominance information.
Aspect ratio	Ratio of the expected width divided by height of the picture content of a programme. Usually 4:3 in SD, 16:9 in HD...
Backup	Action (or result of) copying contents, in the view of being able to recover them, should the originals be corrupted.
Beta SP	Betacam SP, very widespread analogue video tape recording media. Uses MP.
Bit Rot	Random corruptions of one or several bits in a data storage device or medium
Blocking	Artefacts where the coded blocks become visible on the pictures
BNC	A widespread cable connector for transmitting video (analogue, SDI, HD-SDI...)
Born-Digital	Content that has gone through no analogue storage step
Carrier	Physical media, holding video, or files.
Cloud storage	A model where data and files are not stored on a physical medium, but where storage is made seamlessly on a set of interconnected systems in distributed (possibly third-parties) locations, in the view of ensuring a greater availability and robustness to individual component failures.
Codec	“Coder-Decoder”; device or computer program capable of encoding or decoding a digital data stream or signal.
Colour spaces	System by which the colours in the contents are represented. A colour is represented by three coordinates in a colour space, but if an error is made changing from one colour space to another, or if the wrong colour space is specified, representation of colours will be affected. (a short list of colour spaces[5] : RGB, R’G’B’, CYMK, CIE XYZ, L*u*v*, L*a*b*, Y 'C_BC_R, YcCbcCrc, Y ’P_BP_R, Y'UV, Y'IQ…)
Container	See Wrapper
CRC	Cyclic Redundancy Checks (CRC). Codes that allow detecting errors in a bit stream.
DCT	Discrete Cosine Transform. Mathematical transformation of data, used in lossy compression.

[5] http://www.poynton.com/PDFs/ColorFAQ.pdf

Digibeta	Digital Betacam, a wide-spread 4:2:2 digital video recording technology on tape, uses Metal Particle ½ Inch tape in cassettes.
Digitisation	Action of digitising analogue contents into a Digital video tape or into one or several data files.
DNS	Domain Name Servers
DNX	Digital Nonlinear Extensible - High-Definition: Avid de facto standard, approved as SMPTE VC-3. Used in an MXF container, QuickTime also supported. 220 Mbps with a bit depth of 10 or 8 bits, and 145 or 36 Mbps with a bit depth of 8 bits.
Drop-out	Loss of Radio Frequency (RF) signal read by one (or more) of the microscopic magnetic playback heads, reading at high speed (several m/s) the recorded magnetic track, resulting in temporary loss of signal.
DV	The first of the DV digital video tape family, uses intra-frame video compression scheme. Uses 4:1:1 or 4:2:0, 8 bits, at 25Mbps, stored in cassettes, standalone files, of MXF (or others) wrappers.
DVCAM	Professional Sony version of DV
DVCPRO	Professional Panasonic version of DV. Uses MP.
DVCPRO50	Professional Panasonic version of DV, higher quality, 50Mbps
DVCPRO100	HD Panasonic version of DV, 100Mbps
ECC	Error Correcting Codes. Codes that allow detecting and correcting errors (if not too numerous) in a bit stream or file.
Echoes	Artefacts, due to reflections in analogue video signals cables or terrestrial links, resulting in the presence of several weak repetitions of the picture.
File System	An organisation for storing, retrieving, and updating a set of files, on a carrier that is usually a hard disc drive, or a RAID.
Ghosts	See Echoes
GOP	Group Of Pictures: a GOP number higher than 1 (long GOP) implies that some frames/fields (P and B) cannot be reconstructed without using some other frames/fields (I). May improve compression ratios at a given quality, but may make editing more difficult.
HD	High Definition (up to 1920x1080 pixels, up to 50 or 59.94 fields per second)
HDD	(magnetic) Hard Disk Drive
Head-clogging	Transient or permanent loss of signal due to debris accumulated on magnetic heads.
Header	First part of a file, storing metadata and other information, allowing identifying and decoding the rest of the file. A corrupted header may render a file completely unusable.
IMX (D10)	Sony Betacam IMX video format, using MPEG2 4:2:2 intra-frame compression at 30, 40, or 50Mbps. Available as cassettes, but can be exported as MXF files (known as MXF-D10).
LTO	Linear Tape Open: LTO-1 to LTO-6: Series of magnetic tape data cartridges storage technology developed from late 1990s as an open standard alternative to the proprietary magnetic tape formats available at the time. Supported by a number of vendors, has captured the largest share of data tape cartridges market.

MAM	Media Assets Management systems
Mbps	Mega-bits per second: one million bit per second.
Metadata	Data about the data. Information about the contents.
Migration	Moving data, contents, or applications, from one system to another. Migration can involve copying files, but also changing the encodings, the wrappers, the databases, the interfaces....
Matroska	An open source multimedia container/wrapper
ME	Metal Evaporated. Types of tapes with very thin magnetic layers obtained through deposition of metal vapour onto tape. These tapes are considered as very sensitive to humidity and heat, and as more fragile than MP or oxide tapes.
MP	Metal Particle. Tapes where the particles holding the magnetic information are made of very finely ground metallic particles. Allow for a higher storage density than oxide.
MXF	Media Exchange Format. A widely used, but complex, standard for storing as files audiovisual content programmes.
MPEG-2	DCT-based standard for storing as files or bitstreams audiovisual contents. Can support a wide range of intra or inter-frame compression ratios.
NTFS	New Technology File System (Microsoft)
OP	Operational pattern: MXF defines several of these. Most commonly used are OP-atom and OP1a.
OP1a	One of the MXF Operational patterns used to store, in a single file, one video track, with several synchronous audio tracks.
OP-atom	One of the MXF Operational patterns used to store, in a single file, only one track. As a programme is usually composed of at least one video track and one audio track, sort of a hierarchy is usually used to store audio tracks, and other metadata, alongside with the video track(s).
Oxide	Ferric or Chromium oxide: Tapes where the particles holding the magnetic information are made of very finely ground metal oxide particles. Betacam uses oxide.
Overshoot	Artefact from analogue video recordings where oscillations in the analogue video signal become visible as negative replicas around vertical.
P2	A line of solid-state physical media carriers, by Panasonic, in use fir media recording in the field. Can store DV, DVCPRO, DVCPRO25, DVCPRO50, DVCPRO-HD, or AVC-Intra. Uses MXF-OP-atom in a hierarchy, can be accessed as generic storage device.
Pixel	A Picture element. A sample of the picture; may have RGB, YUV, or luminance-only value.
Poisoning	Situation where a number of files are affected by some defect (e.g. wrong parameter), going unnoticed until consequences start to cause real trouble, possibly revealed on the occasion of a migration.
Profile	In this document, the full set of standards, specifications, and parameters that describe up to fine details the encoding for a series of video files. Such profiles can be fully explicit, and published, or implicit (using default or vendor-specific processes or parameters), or even unknown (the profile is not described, by analysis of the file). Changing even one parameter of a profile results in a different profile, or sub-profile. A profile may become a de-facto standard.

ProRes	Apple ProRes 422 is a DCT based intra-frame codec, in 4K, 2K, HD, SD resolutions.
PTS	Presentation Time Stamp: in a coded programme, specifies the specific time in the timeline where the referred chunk of data/video/audio should be presented.
RAID	Redundant Array of Inexpensive Disks: a set of hard disk drives, installed into a physical enclosure, organised so that greater capacity, throughput, and/or robustness to individual component failure is achieved.
Repairing	Correcting a file that has some flaw, e.g. a wrong technical parameter that makes use harder. Correcting even only one parameter over a million files is something that has to be considered carefully...
Resolution	The number of pixels horizontally times the number of pixels vertically (e.g. 720x576, 1920x1080...)
SD	Standard Definition (usually 720x576, 50i (Europe), or 720x488, 59.94i (US, JP...))
SDI	Serial Digital Interface: Standard for transmitting SD video (plus possibly audio and ancillary data) at 270Mbps over a cable (300 m typical).
S-DLT	Super-Digital Linear Tape: a magnetic tape data storage technology. One-vendor. Superseded by LTO.
Streaking	An exponential decay memory effect, potentially affecting analogue transmitted video, where contrasted objects are followed (on the right hand side) by a trail, and look flat and transparent.
U-Matic	Sony ¾ Inch analogue video tape format, 1971 - ~1995
VTR	Video Tape Recorder (and player).
Wrapper	A standard describing how different media elements and metadata coexist in a computer file (e.g. MXF, QuickTime, Matroska, WAV, MPEG-4…).
WAV	A Wrapper for audio contents.
XDCAM	Sony line of recording media products, based on Blu-Ray, within shell. Uses MXF-OP1a, can store DVCAM, IMX, XDCAM HD, XDCAM EX and XDCAM HD422 can be accessed as generic storage device.
ZFS	File system and logical volume manager from Sun Microsystems (originally “Zettabyte File System”)

Table 6 : Glossary