



---

**BBC**

*Research & Development*  
*White Paper*

*WHP 241*

---

*April 2013*

**A Guide to Understanding BBC Archive MXF Files**

**M. Glanville *and* T. Heritage**

*BRITISH BROADCASTING CORPORATION*



## A Guide to Understanding BBC Archive MXF Files

Mark Glanville

Thomas Heritage

### Abstract

The BBC Archive now contains around 15 Petabytes (and growing) of uncompressed files that have been created from videotapes using the BBC R&D Ingex Archive software. The files conform to a custom profile of the sophisticated media file wrapper standard known as *Material eXchange Format* (MXF). The details of the custom profile are fully specified by other BBC R&D White Papers but these documents are not suitable for newcomers to MXF. This White Paper provides an overview of the BBC Archive MXF profile along with a brief introduction to the MXF fundamentals that are required in order to understand it. A diagram of the physical structure of the MXF files is presented – this form of visualisation is found to be very helpful in considering new means of processing the files as well as understanding the consequences of different modes of file corruption.

**Additional key words:** Preservation, Perivale

White Papers are distributed freely on request.  
Authorisation of the Chief Scientist or General Manager is  
required for publication.

©BBC 2013. All rights reserved. Except as provided below, no part of this document may be reproduced in any material form (including photocopying or storing it in any medium by electronic means) without the prior written permission of BBC except in accordance with the provisions of the (UK) Copyright, Designs and Patents Act 1988.

The BBC grants permission to individuals and organisations to make copies of the entire document (including this copyright notice) for their own internal use. No copies of this document may be published, distributed or made available to third parties whether by paper, electronic or other means without the BBC's prior written permission. Where necessary, third parties should be directed to the relevant page on BBC's website at <http://www.bbc.co.uk/rd/pubs/whp> for a copy of this document.

## A Guide to Understanding BBC Archive MXF Files

Mark Glanville                      Thomas Heritage

### 1 Introduction

The BBC Archive contains more than 12 million items including several million television items held on either film or videotape[1]. Migrating the content from physical carriers to files ensures the preservation of content previously held on obsolete carriers, reduces the physical storage space required for the collection, and brings about new opportunities for providing access to the Archive.

In November 2007 the BBC began creating *Material eXchange Format* (MXF) files from television content held on D3 videotapes using the IngeX Archive system developed by BBC R&D (a process known as *ingesting*)[2, 3, 4, 5]. More than 100,000 of these files have now been created which represents around 6 Petabytes of data given that they contain uncompressed video and audio. Since that time the IngeX Archive system, and the MXF files it produces, have been developed and the system is currently in use at the BBC Archive centre in Perivale (West London) for the ingest of Digital Betacam (DigiBeta) videotapes. So far, in total, around 15 Petabytes of BBC Archive MXF files have been produced with IngeX Archive.

MXF is a media file wrapper standard that was developed to harmonise the transfer of video and audio data between devices and organisations – BBC R&D was heavily involved in the design. MXF is widely used in the broadcast industry however the flexibility and broad scope of the MXF standards has led to interoperability issues: MXF files produced by one piece of equipment will not necessarily be understood by another piece of equipment that claims to support MXF files. This has led to *profiles* of MXF being developed for specific applications, most notably the Application Specifications (AS) developed by the Advanced Media Workflow Association (AMWA) [6].

The MXF files produced by the IngeX Archive system follow the BBC Archive MXF file specifications which are described in detail by R&D White Papers[4, 7]. An example file and other information is available from the IngeX Archive webpage[5]. The specifications are designed to provide a relatively small and simple profile of the MXF standards with BBC Archive specific elements. This has the benefit of supporting the BBC specific requirements while minimising complexity and so increasing the likelihood that the files can be understood in the future. However, a drawback of adopting this profile is that the only tools to fully support it are those produced or extended by BBC R&D. A list of relevant tools is included in Appendix B.

Some fundamental MXF concepts are introduced in Section 2, followed by an overview of the BBC Archive MXF profile in Section 3. The latter aims to provide a summary of BBC Archive MXF without assuming additional knowledge of MXF. In addition to the file specifications referenced above it draws on knowledge gained from: *The MXF Book*[8]; IngeX Archive (both its authors and the source code); and examination of the BBC Archive MXF files themselves.

## 2 MXF Basics

As described earlier, MXF is a flexible file wrapper capable of encapsulating many different types of audiovisual material and associated metadata. However, there are fundamental concepts that apply to all MXF files. The most important of these concepts are explained in this Section in the context of BBC Archive MXF.

Further information about how MXF works can be found in *The MXF Book*[8] or by reading the SMPTE standards documents. A list of those most applicable to BBC Archive MXF can be found in Appendix A.

### 2.1 KLV Coding

Everything in MXF is stored physically as *Key Length Value* (KLV) sets. As the name suggests this type of coding has three parts:

- a *key* for identifying the type of set;
- a *length* indicating how long the value is;
- and a *value* containing some information to be stored.

This type of coding allows MXF to be very extensible. If a parser does not understand a particular key then it can simply ignore it, skip ahead in the file by the length given, and read the next key.

#### 2.1.1 Universal Labels (ULs)

All of the keys used in MXF are SMPTE registered Universal Labels (ULs). These are 16-bytes long and start with the bytes 06.0E.2B.34. Reading and comparing 16-byte keys can be resource intensive and they make storage of small values rather inefficient. To avoid doing this there is an option to use 2-byte tags instead of full 16-byte ULs. The mapping between the two is stored in the MXF file's Header and Footer (in the Primer Pack – see Section 3.6). Knowing about the 2-byte tags is useful when looking at the output of `MXFDump` (refer to Appendix B) or examining MXF files in a hex editor. More information about KLV coding can be found in [9] and [8].

### 2.2 Operational Patterns

The MXF design committee defined a set of generalised patterns to describe broad sets of functionality: these are known as *Operational Patterns*. Operational Patterns constrain the complexity of MXF by defining the *structural metadata* to be used. Normally, the higher the number and letter in the name of the Operational Pattern the more complex the MXF file is: so Operational Pattern '1a' (OP1a) is the simplest.

## 3 BBC Archive MXF Files

### 3.1 The Name

This document refers to the MXF profile in question as ‘BBC Archive MXF’. This is not an official term to describe the format but is probably the most descriptive name in use. Other documents may refer to ‘Archive Preservation Project’ MXF files, ‘D3 Preservation Project’ MXF files, or ‘BBC Preservation’ MXF files.

### 3.2 BBC Archive MXF Variations

Before getting into the detail, it should be made clear that there are different types of BBC Archive MXF file. This is a result of the different versions of the software used to create them, IngeX Archive. IngeX Archive has evolved as new features have been added, largely to support the changing set of videotapes to be ingested. There are currently three editions of IngeX Archive and hence three versions of BBC Archive MXF file, each file version adding to the features of the previous version. Unless otherwise stated, this paper describes the version of BBC Archive MXF produced by ‘IngeX Archive HD’ where currently only standard definition (SD) content is considered.

IngeX Archive Edition	White Paper Defining the MXF File Format Used
D3	WHP167 [4]
DigiBeta	WHP233 [7]
HD	[Not yet documented as a White Paper]

‘IngeX Archive D3’ can create BBC Archive MXF files containing 8-bit 4:3 SD uncompressed video data. ‘IngeX Archive DigiBeta’ and ‘IngeX Archive HD’ can create BBC Archive MXF files containing 8-bit or 10-bit SD uncompressed video data with an aspect ratio of 4:3 or 16:9. Clearly ‘IngeX Archive HD’ can also create BBC Archive MXF files containing high definition (HD) video data but that is beyond the scope of this document.

BBC Archive MXF files that were created by any edition of IngeX Archive with SD video may have 4, 5 or 6 audio tracks. Other audio arrangements are supported by ‘IngeX Archive HD’ for HD video data. Audio is sampled at 48 kHz and 24 bits of data are stored for each audio sample in all cases. However, the *metadata* in the MXF file is set to indicate 20-bit quantisation for all versions prior to ‘IngeX Archive HD’. ‘IngeX Archive HD’ sets this metadata to indicate 24-bit quantisation.

### 3.3 Example File Sizes

The use of uncompressed video data results in large files. As an indication of roughly how large, a one hour BBC Archive MXF file is approximately 75 Gigabytes if it contains 8-bit video essence, and approximately 100 Gigabytes if it contains 10-bit video essence. Note that using 10-bit video samples uses one third more storage space than using 8-bit video samples, not one quarter as might be expected. This is because of the byte packing employed.

### 3.4 Essence

Standard definition frame-based essence is stored in BBC Archive MXF files at 25 frames per second.

**Video** There is one track of standard definition video data (720x576, 4:2:2) that is uncompressed and can be either 8-bit UYVY or 10-bit v210. More details of the video coding and packing used can be found in [4] and [7]. The aspect ratio of the video can be 4:3 or 16:9.

**Audio** There can be four, five or six tracks of 24-bit PCM audio data sampled at 48 kHz.

**Timecodes** ‘VITC’ and ‘LTC’ timecode values are ingested from the source videotape if present. The names of the timecode fields refer to their origin, but they are stored in the MXF file as *data* rather than being embedded in the video or audio tracks (although there may also be timecode stored in this way). In addition there is a continuous ‘Control’ timecode that begins at a value of 00:00:00:00 and runs for the duration of the file.

### 3.5 Metadata

#### 3.5.1 Infax

Until recently Infax was the BBC Archive’s metadata database. Its name is used here to refer to a data structure supported by that database, which is recreated in the BBC Archive MXF files. There are two sets of Infax metadata stored in each file: one for the source videotape that the programme was ingested from; and one to cover both the production of the MXF file and the details of the LTO data tape that it was originally stored on. The fields in the Infax metadata are as follows:

- Format
- Programme title
- Episode title
- TX date
- Magazine prefix
- Programme number
- Production code
- Spool status
- Stock date
- Spool descriptor
- Memo
- Programme duration
- Spool number
- Accession number
- Catalogue detail
- Item number

#### 3.5.2 Flags

Some metadata is generated by analysis of the essence during ingest and by the ingest equipment’s self-diagnostic systems. This is stored in the MXF files in the form of *flags* linked to the relevant frames. Unfortunately there is no mechanism to record whether or not a particular detector / flag-generator was running when the file was created. If there are no flags of a particular type this means that the relevant condition was not detected. This may be because the condition did not occur (as far as the detector was aware) *or* that the detector for that condition was not running when the MXF file was created. The different types of flags are:

**PSE Failure** The Harding Flash and Pattern Analyser is used to analyse video for flashes and patterns likely to induce photosensitive epileptic seizures. A PSE Failure flag will include the details of the types of signals that caused the failure: red flashing, spatial flashing or luminance flashing. The details of these failures are explained in [4]. An ‘extended’ failure warning is also possible if flashing persists for more than 5 seconds, but below the threshold for triggering a warning of red or spatial flashing.

**VTR Error** During playback the VTR is continually interrogated by IngeX Archive for its current error level. If this is non-zero an error may have occurred while reading the videotape and this event is stored as a VTR Error flag. Separate values are recorded in the VTR Error flags for video and audio playback errors.

**Timecode Break** When the MXF file is created by IngeX Archive the ‘LTC’ and ‘VITC’ timecode values are analysed to ensure they are increasing monotonically. If this is not the case then a Timecode Break flag is set. Another flag will be set if the timecode returns to monotony. The flags are set independently for LTC and VITC timecodes.

**DigiBeta Dropout** Software to detect the types of videotape dropout commonly observed with Digital Betacam (DigiBeta) was developed in R&D in 2007. This software is run as part of the ingest process and flags are added to the files if the detector reports a probability above a defined threshold.

### 3.5.3 Other Metadata

Various other metadata is also stored in the MXF file Header, and repeated in the Footer. For example:

- Information is stored about the software used to create the MXF file.
- The ‘Network Locator’ is the original file name of (and optionally the path to) the file. Obviously this is not updated when the file is copied, moved or renamed using the operating system’s file management tools, but it can be updated using specialist MXF software.
- The MXF file uses ‘Essence Descriptors’ to hold details about the essence in the file, such as the sample rate, bit depth and aspect ratio.

### 3.6 Physical Structure

BBC Archive MXF files are made up of three parts: Header, Body and Footer. The Header and Footer contain metadata and structural information about the MXF file, and the Body is comprised of ‘Content Packages’. There can be zero or more Content Packages, one per frame. Each Content Package contains the video, audio and timecode data for that frame, along with an optional array of 32-bit Cyclic Redundancy Codes (CRC-32s). CRC-32s are checksums that can be used to ensure the integrity of the video and audio data (with one CRC-32 for each video and audio track).

Figure 1 is a graphical representation of how different parts of the MXF file are stored physically. A list is given here of important details which are either not included on the diagram or which benefit from being stated explicitly. Refer to Section 3.7 for details of the *logical* structure of the file contents and an explanation of the different types of ‘Packages’:

- The duration of the file (the number of frames) is stored in various places throughout the file. The Tape Source Package contains an additional duration value that is fixed at 120 hours and bears no relation to either the duration of the source videotape or the MXF file. This value should be ignored.
- The Material Package and the Tape Source Package contain a name derived from the ‘Spool Number’ stored in the Source Infax Metadata.
- There are 8 ‘creation’/‘modification’ timestamps stored throughout the file, all of which are equal and set to the original creation date of the file.
- The ‘Fill’ element<sup>1</sup> at the end of the Header allows the Header to be edited without needing to rewrite the whole file. It ensures that the Body starts at byte 32768 (0x8000).
- There are small ‘Fill’ elements immediately after each Infax element.
- The ultimate value of the Header’s *metadata status* property is ‘Open, Complete’ and that of the Footer is ‘Closed, Complete’ (note that the Header always remains ‘Open’). These ultimate values are reached: only after the MXF file is updated for writing to LTO tape under ‘Ingex Archive D3’; as soon as ingest has been completed for subsequent versions of Ingex Archive.
- The Footer is almost identical to the Header apart from:
  - The details of the various flags are only stored in the Footer (flag *details* are only present if the flag count is greater than zero).
  - There is no ‘Fill’ at the end of the Footer.
  - The Footer ends with a ‘Random Index Pack’, detailing the locations of the Header Partition and the Footer Partition.
  - The ‘Primer Pack’ in the Footer contains additional 2-byte tags (see Section 2.1.1) for the flag details.
  - The Header and Footer Index Tables are identical for BBC Archive MXF files created with ‘Ingex Archive HD’. However, those created by previous versions of Ingex Archive had a value of zero for the ‘File Duration’ field in the Header Index Table and the correct duration in the Footer Index Table.
  - The Header and Footer Metadata Sets are identical except for a small amount of MXF structure.

---

<sup>1</sup>A ‘Fill’ element is an empty metadata item comprised of null or meaningless data



### 3.7 Logical Structure

The *logical* structure of the contents of a BBC Archive MXF file is different to the physical layout – BBC Archive MXF files can be represented *logically* as three ‘packages’:

- Material Package;
- File Source Package;
- Tape Source Package.

The File Source Package describes the essence as it is stored in the file. It contains descriptive metadata about the file itself and how it was made (i.e. about the ingest process). The Tape Source Package represents the videotape from which the MXF file was ingested and contains the tape’s metadata. Finally, the Material Package references the File Source Package and defines how the essence should be played back from the MXF file.

BBC Archive MXF uses the simplest Operational Pattern: OP1a (see Section 2.2). This means that there is only one Material Package in the file and it is the same duration as the File Source Package. So, the MXF file contains a single, continuously playable item of audio-visual content.

A detailed diagram of the logical structure of a BBC Archive MXF file produced by ‘Ingest Archive D3’ is available from [5].

#### 3.7.1 Unique Material IDs (UMIDs)

Each package is identified by a globally unique 32-byte *Unique Material ID* (UMID). Each BBC Archive MXF file produced by Ingest Archive has a unique UMID generated for each of the three packages. All three values are stored in the Ingest Archive database. This is the case even if a videotape is ingested several times and so several MXF files are produced – the expectation may be that the files would all have Tape Source Packages with identical UMID values (in order to identify that they all originated from the same videotape) but this is not the case. This means that in practice the File Source Package and Tape Source Package UMIDs can be ignored but the Material Package UMID is useful as a globally unique identifier for the MXF file.

## References

- [1] Sarah Hayes. *The new BBC Archive Centre at Perivale*. October 2011. [http://www.live.bbc.co.uk/blogs/aboutthebbc/posts/the\\_new\\_bbc\\_archive\\_centre\\_at](http://www.live.bbc.co.uk/blogs/aboutthebbc/posts/the_new_bbc_archive_centre_at).
- [2] Adrian Williams. *Safeguarding the BBC's Archive*. August 2010. [http://www.bbc.co.uk/blogs/bbcinternet/2010/08/safeguarding\\_the\\_bbc\\_archive.html](http://www.bbc.co.uk/blogs/bbcinternet/2010/08/safeguarding_the_bbc_archive.html).
- [3] Stuart Cunningham & Philip de Nier. *File-based Production: Making It Work In Practice*. White Paper 155, BBC Research & Development, September 2007. <http://www.bbc.co.uk/rd/pubs/whp/whp155.shtml>.
- [4] Philip de Nier & Phil Tudor. *D3 Preservation File Format*. White Paper 167, BBC Research & Development, July 2008. <http://www.bbc.co.uk/rd/pubs/whp/whp167.shtml>.
- [5] *Ingex Archive*. <http://ingex.sourceforge.net/archive/>. Accessed: 14 January 2013.
- [6] *AMWA Application Specifications*. [http://www.amwa.tv/projects/application\\_specifications.shtml](http://www.amwa.tv/projects/application_specifications.shtml). Accessed: 31 January 2013.
- [7] Philip de Nier. *Archive Preservation File Format: DigiBeta system*. White Paper 233, BBC Research & Development, September 2012. <http://www.bbc.co.uk/rd/pubs/whp/whp233.shtml>.
- [8] Nick Wells, editor. *The MXF Book*. Focal Press, 2006.
- [9] Omneon Inc. *Encoding Data into MXF Files: BER and KLV encoding*. March 2010. <http://www.amwa.tv/downloads/whitepapers/encodingtoMXF.pdf>.

## Appendix A List of SMPTE Standards Documents Pertinent to BBC Archive MXF

This list is not necessarily exhaustive and obviously these documents will refer to others. The MXF Book[8] *really is* a better place to start!

- SMPTE ST 377-1 : MXF File Format
- SMPTE ST 378 : Operational Pattern 1A
- SMPTE ST 379 : Generic Container
- SMPTE ST 380 : Descriptive Metadata Scheme- 1
- SMPTE ST 382 : AES and BWF Audio
- SMPTE ST 384 : Uncompressed Pictures
- SMPTE ST 330 : Universal Material Identifier (UMID)
- SMPTE ST 336 : KLV Coding
- SMPTE ST 331 : Content Package Elements and Metadata

## Appendix B Software Support for BBC Archive MXF Files

This list is believed to be correct at the time of writing but is not definitive. Only a small selection of the tools provided by each project or software library are listed here.

### Ingex

As well as Ingex Archive, the Ingex suite of tools includes **player**: a software tool for playing (among other things) MXF files. It is able to read and display the metadata, metadata flags and timecodes stored within BBC Archive MXF files.

URL: <http://ingex.sourceforge.net/>

### bmX

The bmX project is home to the LibMXF library, which is used by Ingex Archive to create the files on ingest from videotape. LibMXF++ provides a C++ interface to the functions in LibMXF. Furthermore, there are many C++ classes in the Ingex Archive source code that provide yet further abstraction and hence simpler use of LibMXF. Also provided by the bmX project are: **archive\_mxf\_info** and **mxf2raw**, which can read BBC Archive MXF files and output a summary of their essence and metadata; **MXFDump**, which can parse the contents of MXF files and produce a detailed text output for the entire file – useful for examining the physical structure of MXF files.

URL: <http://sourceforge.net/p/bmxlib/>

### MXFLib

A free, open source MXF library, similar to LibMXF. It doesn't have full support for BBC Archive MXF files, but can read the essence of the files and its **mxfDump** tool can be used to parse their contents (it is similar to **MXFDump** mentioned above but is more suited to examining the logical structure of MXF files).

URL: <http://sourceforge.net/projects/mxflib/>

### libavformat

Originating in the *FFmpeg* project, libavformat is a software library capable of reading the essence of BBC Archive MXF files. It is used by numerous free and open source software tools including VLC, FFmpeg and FFmpeg.

URL: <http://www.ffmpeg.org/>