We have established that low level data corruptions exist and that they have several origins. The error rates are at the $10^{-7}$ level, but with complicated patterns.
Securing the past for the future
1: FILENAMES

BKD029986.mxf
1: FILENAMES

BKD029986_79054025255fb1a26e4bc422aef54eb4.mxf
2: MODULARITY
3: SOX : FFT
3: SOX : FFT
4: BARCODES
5: PHYSICAL SCANNING
PHYSICAL SCANNING
6: VTR HEAD ERRORS
6: VTR HEAD ERRORS
6: VTR HEAD ERRORS
6: VTR HEAD ERRORS
7: FILE TRIMMING
8: LONG TERM
8: LONG TERM

_Approach_
The digitisation of selected videotapes will be carried out at BBC Archives’ Odyssey Park facility in South Ruislip using workflow systems built in-house by BBC Archives. These use Telesystem Lightspeed units to create the uncompressed master files, and FFmpeg to create IMX50 files from these, which are then re-wrapped to ASI1 DPP files using bmxtranswap.

BBC Archives have committed to a ‘touch once’ approach to the digitisation of these assets. This means each videotape digitisation should produce:

- A preservation quality digital copy of the videotape as Uncompressed 10-bit YUV 720x576 (v210) Quicktime (MOV) file with four 48kHz 24-bit uncompressed audio tracks. These un压缩ed files are only stored on these LTO tapes.

- A broadcast quality digital copy of the videotape as an ASI1 DPP MXF (720x608 IMX/D10 50Mbps with four channel 48kHz 24-bit audio) which will be imported to the BBC’s Digital Archive (VizRT’s Viz One MAM) resulting in these files being stored to a 2+1 storage level.

- Accompanying digital images of the physical asset, its packaging and any related VTR documentation that was enclosed within.

_Clipping_
The ASI1 DPP MXF file will be clipped where possible to remove trailing black (or grey) and silence (with a 15 second handle) in order to minimise wastage of storage space in the BBC Digital Archive.
for the given track. This checksum was produced by mxf2raw during extraction of the MXF file from LTO tape. The values of these checksums are not stored inside the MXF file.

This XML document was produced by analysing D__37150401-LTA00089207.mxf.full_mxf_analysis.xml.

D__37150401-LTA00089207.mxf.full_mxf_analysis.xml

This XML file contains the full details produced by using mxf2raw to analyse the MXF file. Also refer to the summarised version of this XML file and its explanation.

D__37150401-LTA00089207.mxf.md5

This file contains the MD5 checksum for the entire MXF file. This checksum was produced by mxf2raw during extraction of the MXF file from LTO tape. This means that it was produced by the same process that was also fully reading (and partially verifying) the MXF file. This checksum also appears in the XML files included in this collection.

D__37150401-LTA00089207.mxf.sha1

This file contains the SHA-1 checksum for the entire MXF file. This checksum was produced during the extraction of the MXF file.
Figure 5 Preface and content storage packages

The Preface set contains two BBC extension items, D3P D3 Error Count and D3P PSE Failure Count, defined in Table 1. These count values are useful when reading a file from an LTO tape because the series of D3 VTR replay errors and PSE analysis failures are recorded in the footer section and are only available once the whole file has been transferred from the tape.

<table>
<thead>
<tr>
<th>Name</th>
<th>Meaning</th>
<th>Type</th>
<th>UL</th>
</tr>
</thead>
<tbody>
<tr>
<td>D3P D3 Error Count</td>
<td>Total number of D3 VTR replay errors</td>
<td>UInt32</td>
<td>06.0e.2b.34.01.01.01.01.0d.04.01.01.40.01.01.00</td>
</tr>
<tr>
<td>D3P PSE Failure Count</td>
<td>Total number of PSE analysis failures</td>
<td>UInt32</td>
<td>06.0e.2b.34.01.01.01.01.0d.04.01.01.40.01.02.00</td>
</tr>
</tbody>
</table>
A Guide to Understanding BBC Archive MXF Files

M. Glanville and T. Heritage

April 2013
const MXFMCA::LabelIndex *mca_label_index = reader->GetMCALabelIndex();
map<uint32_t, vector<AudioChannelLabelSubDescriptor>> c_labels;
map<UUID, SoundfieldGroupLabelSubDescriptor> sg_labels;
map<UUID, GroupOfSoundfieldGroupsLabelSubDescriptor> gosg_labels;

size_t i;
for (i = 0; i < sound_info->mca_labels.size(); i++) {
  AudioChannelLabelSubDescriptor *c_label = sound_info->mca_labels[i];
  BMX_CHECK(c_label->haveMCACChannelID());
  c_labels[c_label->getMCACChannelID()].push_back(c_label);

  if (c_label->haveSoundfieldGroupLinkID()) {
    UUID link_id = c_label->getSoundfieldGroupLinkID();
    MCALabelSubDescriptor *label = mca_label_index->FindLabel(link_id);
    SoundfieldGroupLabelSubDescriptor *sg_label = dynamic_cast<SoundfieldGroupLabelSubDescriptor*>(label);
    BMX_CHECK(sg_label);
    sg_labels[link_id] = sg_label;
  }

  if (sg_label->haveGroupOfSoundfieldGroupsLinkID()) {
    vector<UUID> link_ids = sg_label->getGroupOfSoundfieldGroupsLinkID();
    size_t k;
    for (k = 0; k < link_ids.size(); k++) {
      UUID &link_id = link_ids[k];
      MCALabelSubDescriptor *label = mca_label_index->FindLabel(link_id);
      GroupOfSoundfieldGroupsLabelSubDescriptor *gosg_label = dynamic_cast<GroupOfSoundfieldGroupsLabelSubDescriptor*>(label);
      BMX_CHECK(gosg_label);
      gosg_labels[link_id] = gosg_label;
    }
  }
}
<table>
<thead>
<tr>
<th>tankenumber</th>
<th>bbc.firsttxdate</th>
<th>bbc.firsttxservice</th>
<th>asset.materialtype</th>
<th>asset.siteidentity</th>
<th>bbc.programmenumber</th>
<th>bbc.episodetitle</th>
</tr>
</thead>
<tbody>
<tr>
<td>0301</td>
<td>31/01/1995</td>
<td>00:00</td>
<td>prov</td>
<td>LLCF545T/71</td>
<td>LLCF545T/71</td>
<td>JIM TAVARE PICTURES</td>
</tr>
<tr>
<td>0200</td>
<td>04/03/1992</td>
<td>00:00</td>
<td>prov</td>
<td>LTFC143S/71</td>
<td>LTFC143S/72</td>
<td>PRIMETIME</td>
</tr>
<tr>
<td>0105</td>
<td>16/10/1993</td>
<td>00:00</td>
<td>prov</td>
<td>LMAA855A/71</td>
<td>LMAA855A/71</td>
<td>DANNY BAKER AFTERNOON</td>
</tr>
<tr>
<td>0101</td>
<td>29/10/1994</td>
<td>00:00</td>
<td>prov</td>
<td>LMBB717L/72</td>
<td>LMBB717L/72</td>
<td>THE LAST WORD</td>
</tr>
<tr>
<td>0030</td>
<td>08/11/1993</td>
<td>00:00</td>
<td>prov</td>
<td>LMAV793T/01</td>
<td>LMAV793T/01</td>
<td>INSIDE VICTOR LECGRAY</td>
</tr>
<tr>
<td>0032</td>
<td>30/12/1992</td>
<td>00:00</td>
<td>prov</td>
<td>LMAV740F/71</td>
<td>LMAV740F/71</td>
<td>TALKING REMBRANDTS</td>
</tr>
<tr>
<td>0034</td>
<td>06/04/1994</td>
<td>00:00</td>
<td>prov</td>
<td>LMAA192X/71</td>
<td>LMAA192X/71</td>
<td>YOUNG MUSICIAN ON TOUR</td>
</tr>
<tr>
<td>0036</td>
<td>07/02/1984</td>
<td>00:00</td>
<td>prov</td>
<td>LCHD311F/01</td>
<td>LCHD311F/01</td>
<td>GRANGE HILL'S 11</td>
</tr>
<tr>
<td>0028</td>
<td>24/11/1978</td>
<td>00:00</td>
<td>prov</td>
<td>LCHK040A/71</td>
<td>LCHK040A/71</td>
<td>JACKANORY: THE DRIP</td>
</tr>
<tr>
<td>0030</td>
<td>25/01/1982</td>
<td>00:00</td>
<td>prov</td>
<td>LCHJ144Y/71</td>
<td>LCHJ144Y/71</td>
<td>PLAYHOUSE: THE BLUE MAN</td>
</tr>
<tr>
<td>0032</td>
<td>07/01/1980</td>
<td>00:00</td>
<td>prov</td>
<td>LCHJ043D/71</td>
<td>LCHJ043D/71</td>
<td>PLAYHOUSE: THE SADIST</td>
</tr>
<tr>
<td>0034</td>
<td>30/08/1981</td>
<td>00:00</td>
<td>prov</td>
<td>LCHC127E/71</td>
<td>LCHC127E/71</td>
<td>SWAP SHOP'S SUMMER</td>
</tr>
<tr>
<td>0036</td>
<td>18/01/1982</td>
<td>00:00</td>
<td>prov</td>
<td>LCHJ142K/72</td>
<td>LCHJ142K/72</td>
<td>PLAYHOUSE: THE UNDERLIFER</td>
</tr>
<tr>
<td>0038</td>
<td>12/04/1985</td>
<td>00:00</td>
<td>prov</td>
<td>LCHJ501X/72</td>
<td>LCHJ501X/72</td>
<td>WHIZZ: IN THE SHARK</td>
</tr>
<tr>
<td>0030</td>
<td>20/07/1983</td>
<td>00:00</td>
<td>prov</td>
<td>LCHB0093X/71</td>
<td>LCHB0093X/71</td>
<td>BLUE PETER GOES TO THE SEAS</td>
</tr>
<tr>
<td>0056</td>
<td>20/01/1982</td>
<td>00:00</td>
<td>prov</td>
<td>LCHD235N/71</td>
<td>LCHD235N/71</td>
<td>THE STORY OF THE WOODMAN</td>
</tr>
<tr>
<td>0032</td>
<td>06/04/1980</td>
<td>00:00</td>
<td>prov</td>
<td>LCHC121P/71</td>
<td>LCHC121P/71</td>
<td>MULTI-COLOURED SKIES</td>
</tr>
<tr>
<td>0035</td>
<td>05/01/1983</td>
<td>00:00</td>
<td>prov</td>
<td>LCHD045B/71</td>
<td>LCHD045B/71</td>
<td>CAPTAIN ZEP: SPACE</td>
</tr>
<tr>
<td>0034</td>
<td>17/01/1979</td>
<td>00:00</td>
<td>prov</td>
<td>LCHD003Y/71</td>
<td>LCHD003Y/71</td>
<td>THE STRANGE AFFAIRS</td>
</tr>
<tr>
<td>0054</td>
<td>30/05/1983</td>
<td>00:00</td>
<td>prov</td>
<td>LCHC130L/71</td>
<td>LCHC130L/71</td>
<td>MIKE READ'S REPETITION</td>
</tr>
<tr>
<td>0056</td>
<td>02/05/1983</td>
<td>00:00</td>
<td>prov</td>
<td>LCHC129S/71</td>
<td>LCHC129S/71</td>
<td>MIKE READ'S REPETITION</td>
</tr>
<tr>
<td>0036</td>
<td>08/04/1974</td>
<td>00:00</td>
<td>prov</td>
<td>LCH2567S/01</td>
<td>LCH2567S/01</td>
<td>PLAY SCHOOL</td>
</tr>
<tr>
<td>0032</td>
<td>05/06/1972</td>
<td>00:00</td>
<td>prov</td>
<td>LCH2525P/01</td>
<td>LCH2525P/01</td>
<td>PLAY SCHOOL</td>
</tr>
<tr>
<td>0036</td>
<td>25/05/1972</td>
<td>00:00</td>
<td>prov</td>
<td>LCH9228Y/01</td>
<td>LCH9228Y/01</td>
<td>BLUE PETER</td>
</tr>
<tr>
<td>0032</td>
<td>14/02/1974</td>
<td>00:00</td>
<td>prov</td>
<td>LCH9339K/01</td>
<td>LCH9339K/01</td>
<td>BLUE PETER</td>
</tr>
</tbody>
</table>
BBC Archives

Securing the past for the future

steve.daly@bbc.co.uk
A: TIMECODE
B: FILE FORMATS

Preservation: V210 Uncompressed Quicktime

Access: AS11 DPP MXF
C: FACTORY - LEAN

Prepare  Diagnose  Design  Plan  Implement  Sustain

Processes

Think and feel

Manage and Measure
D: OPEN SOURCE TOOLS

ffmpeg / ffmbc
bmxlib (bmxtranswrap & mxf2raw)
mediainfo
sox
python
postgresql / mysql