

Unerwartete Obsoleszenzen in dateibasierten AV-Archiven

Jörg Houpert

Mit der Einführung von dateibasierten Workflows in Medienorganisationen haben auch vielfältige audiovisuelle Dateiformate ihren Einzug in AV-Archive gefunden. Produkte, die Mediendateien verarbeiten, haben sich in kurzen Zyklen dynamisch weiterentwickelt und haben dabei auch immer besser die Vorgaben der Medienstandards befolgt. In den Videoarchiven lagern nun jedoch Mediendateien mit den ‚Kinderkrankheiten‘ aus zehn Jahren Lernerfahrungen der Hersteller von dateibasierten Medienprodukten. Auch die Medienformatstandards selbst mussten einen Reifungsprozess durchlaufen, mit neuen - nicht immer rückwärtskompatiblen Versionen - und immer detaillierteren Beschreibungen, wie die Standards zu interpretieren sind. Interessant ist auch, wie sich die Messgeräte in ihrer Fähigkeit, die Unzulänglichkeiten bei AV-Dateiformaten zu erkennen, weiterentwickelt haben.

Wird die erste Dekade des Videodateizeitalters womöglich langfristig die gleichen Lücken aufweisen, wie die erste Dekade des Stummfilms, oder die erste Dekade der Wachswalzen-Tonaufzeichnung? Werden technologisch begabte Archivare in Zukunft die Bits und Bytes der Medienformate variieren, um doch noch auf einzelne Inhalte zugreifen zu können, so wie heute Experten die richtige Abspielgeschwindigkeit von frühen Schellackplatten rein nach Gehör versuchen zu rekonstruieren? Oder wird einer der nächsten DOT.com Start-up Unternehmen einen Algorithmus entwickeln, der alle Medienformatfehler vollautomatisch reparieren kann?

Der Autor arbeitet seit fast 20 Jahren in der Standardisierung und

Qualitätssicherung von Medienformaten. Einerseits im Rahmen der Europäischen Fernsehunion (EBU), andererseits ist er auch seit vielen Jahren in der internationalen Standardisierung als Professional Member innerhalb der Society of Motion Picture & Television Engineers (SMPTE) tätig. Hinzu kommen Arbeiten innerhalb der Advanced Media Workflow Association (AMWA). Standardisierungsarbeiten erfolgen auch im Technical Committee der International Association of Sound and Audiovisual Archives (IASA) und der Preservation & Migration Commission der International Federation of Television Archives (FIAT/IFTA).

Schauen wir uns im Folgenden etwas genauer an, worin die Abspielprobleme der Mediendateien im Fernseharchiv bestehen. Im Fernsehbetrieb wird fast ausschließlich das Dateiformat MXF (Material Exchange Format) verwendet. MXF ist ein offener, internationaler Standard und wurde 2004 von der SMPTE eingeführt.

MXF ist ein sehr flexibler Standard. Es handelt sich hierbei um ein sogenanntes Containerformat, welches bereits am Markt verfügbare Medienformate transportieren kann. Hierhinein wurden verschiedenste bestehende herstellerspezifische Videokodierungsformate integriert.

Diese hohe Flexibilität mit ihren vielen Freiheitsgraden führt einerseits zu einer sehr hohen Hersteller- und Anwenderakzeptanz, andererseits bringen die vielen erlaubten Freiheitsgrade aber auch ihre Probleme mit sich. Diese Probleme können in zwei Gruppen aufgeteilt werden; erstens eine unzurei-



Jörg Houpert
Technische Leitung
Cube-Tec International GmbH
T: 0421 20144 0

chende Standardkonformität und/oder zweitens eine unzureichend Interoperabilität (sprich die Fähigkeit eine von Produkt A erzeugte MXF-Datei in einem Produkt B fehlerfrei interpretieren und weiterverarbeiten zu können). Wieso besteht überhaupt ein Unterschied zwischen Standardkonformität und Interoperabilität? Sollte nicht eine Standardkonformität automatisch eine Interoperabilität beinhalten? Ist das nicht sogar der Hauptgrund weswegen es Standards überhaupt gibt?

In einer idealen Welt gibt es tatsächlich keine Interoperabilitätsprobleme solange ein Standard von beiden Produkten eingehalten wird. In der Praxis werden Standards inzwischen jedoch so umfangreich und mächtig, dass Hersteller diese nicht mehr vollständig unterstützen (können), sondern nur für ihren Anwendungsbereich relevante Teile daraus. Treffen zwei Produkte aufeinander, die jeweils unterschiedliche Teile der MXF-Standards implementiert haben (und gehen wir davon aus, dass sie dies in einer vollkommen standardkonformem

* Vortragsmanuskript (gehalten auf der Frühjahrstagung des vfm am 26. April 2016)

Weise gemacht haben), so müssen ihre Produkte unter bestimmten Bedingungen trotzdem nicht interoperabel sein. Wir ahnen schon, dass es hierbei leicht zu recht verwickelten Situationen kommen kann.

“Nur ein Drittel aller Rundfunkprodukte erzeugt standardkonforme MXF-Dateien.”

Wenn wir dann weiter betrachten, dass die typischerweise in deutschsprachigen Rundfunk eingesetzten MXF-Produkte, wie es aktuelle Messungen des Institut für Rundfunktechnik (IRT) gezeigt haben (die Messungen erfolgen in jährlichem Abstand innerhalb der MXF Plug-Fests), dass im Durchschnitt nur ein Drittel aller Produkte wirklich standardkonforme MXF-Dateien erzeugt, so wird nachvollziehbar, warum tausende von Fehlerquellen bei einem einzigen MXF-Format möglich sind. Messungen bei Cube-Tec bestätigen, dass diese große Vielzahl von Formatvariation auch in der Praxis vorkommt. Aus meiner Sicht besteht der einzige Grund, warum auf der Anwenderseite kein lauter Aufschrei entsteht, darin, dass fast alle Anwender sich einfach nicht vorstellen können, welche komplexen Fehlerbilder ihre Dateien in Wirklichkeit in sich tragen. Es existieren jedoch noch weitere Gründe, warum in der Praxis der Programmaustausch trotz dieser Sachlage noch erstaunlich gut funktioniert. Rundfunkanwender testen vor einer Inbetriebnahme alle Produktionswege durch und verwenden neue Produktversionen erst nachdem sie in ihren Testsystemen das Zusammenspiel gründlich überprüft haben. Der zweite Grund ist, die Hersteller haben alle begonnen fehlerhafte MXF-Dateien intern mit ihren eigenen „propriären“ Methoden zu korrigieren. Was auf den ersten Blick wie eine

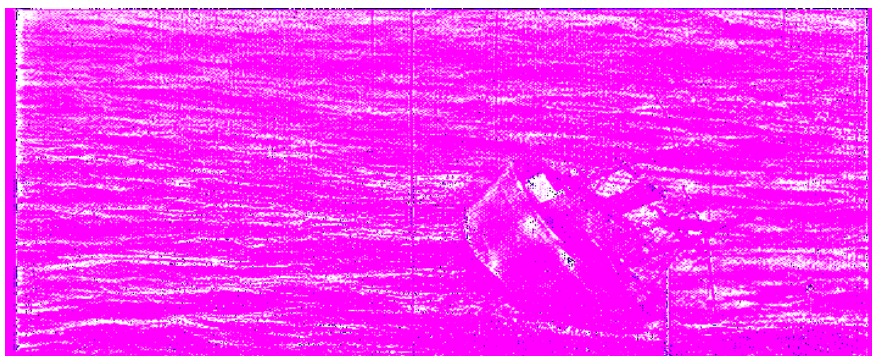


Abb. 1 und 2

lobenswerte Reaktion der Hersteller erscheint, führt in der Praxis jedoch zu noch komplexeren Fehlersituationen, da das Verhalten dieser „propriären“ Fehlerkorrekturmethoden sich jedem externen Einblick entzieht. Zusätzlich verschleiert es Fehlersituationen und untergräbt damit für den Anwender eine leichte Überprüfbarkeit der Standardkonformität.

Im Rahmen des Europäischen Forschungsprojektes *DAVID - Digital AV Media Damage Prevention and Repair*¹ hat sich der Autor zweieinhalb Jahre mit Fragestellungen und Möglichkeiten zur Reparatur von Mediendateiformaten beschäftigt.

Aus den Forschungsergebnissen lassen sich die Risiken für zukünftige Format-Obsoleszenzen einigermaßen seriös abschätzen und es lassen sich klare Handlungsempfehlungen ableiten, wie diese Risiken zu minimieren sind.

Das DAVID Forschungs- und Entwicklungs-Konsortium bestand aus Joanneum Research (Österreich) und IT Innovation (England)

als die Partner aus der Forschung. Cube-Tec International, (Deutschland) und HS-ART (Österreich) waren als Forschungs- und Kommerzialisierungspartner aktiv. Als Archivpartner haben das Projekt begleitet: das Institut National de l'Audiovisuel, INA (Frankreich) und der Österreichischer Rundfunk. Die letzten beiden Partner haben auch die Projektergebnisse in praktischen Tests auf Archivtauglichkeit überprüft. Das Projekt hatte eine Laufzeit von 30 Monaten und lief vom Dezember 2012 bis Mai 2015.

Im DAVID-Projekt wurden auch reine Archivformate wie JPEG 2000 als verlustfreies Speicherformat für die Langzeitbewahrung untersucht. JPEG 2000 wird bei der INA als Archivformat verwendet. Hier ein einzelnes Bild aus einem Archivfilm der INA.

Die obere Abbildung zeigt wie das Bild bei der Archivierung im Jahre 2012 ausgesehen hatte. Die untere Abbildung zeigt, wie das gleiche Bild aussah als es im Frühjahr 2014 wieder aus dem Archiv zurückgelesen wurde. Aus der Checksumme der Archivdatei war ersichtlich, dass das File während der Archivierung



Digital AV Media Damage
Prevention and Repair



1) <http://david-preservation.eu/>

nicht verändert wurde und auch die Backup-Kopie zeigte das gleiche Verhalten. Nachdem der DAVID Projektpartner IT Innovation den Fall näher untersucht hatte, stellte sich heraus, dass die Archiveinspielung mit einer weitverbreiteten Freeware Software erfolgte (ffmpeg). Das Auslesen erfolgte ebenfalls mit der gleichen Software, in der zwischenzeitlich aktuellen Version. Was war geschehen? Nach einer genaueren Untersuchung zeigte sich, dass keine der beiden verwendeten Versionen an dieser Stelle fehlerhaft war, sondern, dass ein bestimmtes Merkmal was den zu verwendenden Farbraum definiert in der gewählten Formatvariante bei JPEG 2000 im Standard nicht definiert war und so von den Entwicklern auf gut Glück gesetzt wurde. Scheinbar änderte sich die Entwicklermeinung wie das Glück wirklich aussehen sollte zwischen beiden Versionen. (In einem unserer öffentlich zugänglichen DAVID Forschungsberichten haben wir die Fakten technisch beschrieben (siehe: <http://dx.doi.org/10.7800/304DAVID22>). Interessant an diesem Beispiel ist aus meiner Sicht das typische Ergebnis, wieder etwas flapsig ausgedrückt. „Keiner hat einen wirklich eindeutigen Fehler gemacht, trotzdem ist die Archivdatei nur mit forensischen aufwändigen Methoden wieder zu rekonstruieren“.

Wer soll diese forensische Fleißarbeit in Zukunft finanzieren? Wie wertvoll muss eine Videodatei sein, dass die Finanzierung dafür freigebenden wird und was passiert zukünftig mit allen Dateien, für die es keine Finanzierung zur forensischen Rekonstruktion gibt?

ORF TV Archiv – das Videotape Migration Projekt

In 2003 startete der ORF ein großes Archivprojekt um etwa 150.000 Stunden von analogem Videoinhalten, die in den Videoformaten Panasonic MII, Ein-Zoll-B-Format und U-matic vorlagen, auf Sony

IMX-Kassetten umzukopieren. Sieben Jahre später entschied der ORF anstatt Sony IMX-Kassetten zu verwenden, die neuen Archivmaster direkt im MXF D-10 Dateiformat zu archivieren. Aber in 2010 hatte der ORF noch keinen Archivmassenspeicher, deshalb wurde ein LTO-4 Tape-Roboter Backup-System als Zwischenspeicher genutzt. Daraufhin wurden mehr als 18.000 Stunden Videomaterial in das MXF D-10 Format umgeschnitten.

Ende 2012 bekam der ORF dann seinen Hi-Res-Archivspeicher und ab dieser Zeit konnten Digitalisate direkt in dem neuen Archivspeicher überführt werden. Nun mussten nur noch die MXF-Backupdateien von den LTO-Zwischenspeicher in den neuen Hi-Res-Archivspeicher überführt werden. Dies erwies sich schwieriger als erwartet, da etwa die Hälfte dieser Dateien vom Archivspeicher wegen fehlerhaften Formaten abgelehnt wurde. Daraufhin suchte man einen Spezialisten und Cube-Tec wurde hinzugezogen. Die Cube-Tec Ingenieure analysierten eine Reihe der defekten MXF-Dateien. Sie konnten die Symptome der auftretenden Fehler nachvollziehen und fanden verschiedenartige Fehler im MXF Container und zufällig verteilte Fehler im kodierten MPEG-Bitstrom. Die gute Nachricht war, das enkodierte Bildsignal war in fast allen Fällen unversehrt. Cube-Tec stellte dem ORF einen speziellen Server zur Dateireparatur zur Verfügung und in weniger als einem Jahr konnten mehr als 9.000 Stunden an defekten MXF Dateien digital korrigiert werden.

Hierbei wurden bis zu 2 Terabyte an MXF-Dateien pro Tag verarbeitet. In diesem Prozess konnten mehr als 95% der defekten MXF-Dateien in vollkommen standardkonforme Dateien gewandelt werden und das ohne Transkodierung – also ohne die Videoqualität durch eine Mehrfachkodierung zu beeinträchtigen.

Die Erfahrungen aus diesem Dateirettungsprojekt haben im Weiteren zu der Entwicklung des

Cube-Tec Produkts MXF Legalizer geführt.

RTP TV Archives – Rádio e Televisão de Portugal in Lissabon

Als letztes praktisches Beispiel folgt ein Einblick ein Archivprojekt beim öffentlich-rechtlichen Rundfunk in Portugal. RTP hatte Abstürze im Playout-System, verursacht durch defekte MXF-Files aus dem Archivsystem. Es wurden um die 30.000 MXF-Dateien gefunden, die Kompatibilitätsprobleme mit dem Playout-System aufwiesen.

Um Inhalte lokal einzuspielen, werden bei RTP Quantel- und Omneon-Server sowie ein OmniBus Automatisierungssystem benutzt. Das Archivsystem integriert ein Front Porch Digital DIVArchive und ein Sony PetaSite Langzeitarchiv.

Während der Vorabtests entdeckten die Ingenieure bei Cube-Tec MXF-Dateien, die mit unzulässigen Abweichungen in der Bitrate enkodiert waren, sowie benutzerdefinierte Datenströme, die nicht korrekt innerhalb der MXF-Dateien ausgerichtet wurden. RTP schaffte daraufhin ein MXF Legalizer System an und korrigierte die defekten Files. Dank der positiven Erfahrung aus diesem Projekt werden inzwischen sämtliche MXF-Dateien, die bei RTP in den Langzeitspeicher eingestellt werden, oder daraus abgerufen werden, über hochverfügbare MXF Legalizer Systeme geroutet.

Erkenntnisse und Handlungsempfehlungen aus dem DAVID-Projekt

Wie lassen sich zukünftige AV-File-Obsoleszenz-Risiken minimieren? Die Handlungsempfehlung ist nicht schwierig. Erstens: Vereinheitlichung des versteckten AV-Formatvariantenwildwuchses. Die meisten Anwender sind sich nicht darüber bewusst, dass sie in Wirklichkeit hunderte von versteckten Format-

variationen in ihren Archiven lagern. Hierbei handelt es sich teilweise um standard-konforme Varianten, häufig jedoch auch um nicht standard-konforme. Die Vereinheitlichung muss ohne verlustbehaftete Re-Encoding Prozesse erfolgen, marktübliche Transcoder-Lösungen sind verlustbehaftet und ungeeignet. Sie können genau wie marktübliche Re-Wrapper-Produkte die vorliegenden Fehler in den defekten Dateien noch verschlimmern. Marktübliche Transcoder und Rewrapper sollten ausschließlich bei fehlerfreien Dateien eingesetzt werden, sie sind keine Reparaturwerkzeuge.

Zweitens: sicherstellen, dass bei der Vereinheitlichung keine neuen 'Fehler' entstehen. Diese Forderung resultiert praktisch gesehen darin, einen Experten in das Projekt mit einzubinden, oder einen eigenen Mitarbeiter entsprechend auszubilden.

Cube-Tec hat sein Verfahren bzw. die damit erzeugten Files durch das IRT zertifizieren lassen. Zusätzlich verwendet der Cube-Tec Prozess eine Fremdüberwachung durch den IRT MXF Analyser als anerkanntes Referenzmessinstrument um jedes einzelne korrigierte Files noch einmal von einem neutralen Dritten gegenprüfen zu lassen. Diese Fremdüberwachung erfolgt im Rahmen des Trusted-Legalizer-Prozesses und ist ein integrierter Teil des Legalizer-Verfahrens.

Die jeweiligen Korrekturmaßnahmen sind in einem Report für jede Datei detailliert und präzise protokolliert. Der Report liegt menschen- und maschinenlesbar vor (PDF und XML), er zeigt die Unterschiede zwischen der fehlerhaften Eingangsdatei und der korrigierten Ausgangsdatei.

Das Verfahren ist inzwischen so ausgereift, dass es geeignet ist viele Petabyte große Archivbestände automatisiert zu korrigieren.



Abbildung 3

Wird es in der Zukunft weniger Medienformatprobleme geben?

Die deutschsprachigen öffentlich-rechtlichen Rundfunkanstalten haben sich dem Interoperabilitätsproblem angenommen und unter der Leitung des IRT neue MXF-Profile spezifiziert. Diese MXF-Profile spezifizieren die im Austausch zwischen den Funkhäuser zu verwendenden MXF-Formate sehr vorbildlich, sie minimieren damit die erlaubten Freiheitsgrade für den zukünftigen Programmaustausch. Hier zeigt sich deutsche Gründlichkeit als weltweites Vorbild.

Wenn die neuen MXF-Profile von allen Beteiligten exakt eingehalten werden, können die Probleme beim Programmaustausch hierdurch zukünftig wirklich weitgehend gelöst werden.

Die neuen ARD-ZDF-MXF-Profile werden natürlich nicht nur als Austauschformat zwischen den Anstalten genutzt werden, sondern nach einer Übergangsfrist auch innerhalb der Funkhäuser.

Damit werden die neuen MXF-Profile auch das neue Archivformat bilden und somit die Formatvielfalt in den Rundfunkarchiven zumindest mittelfristig eindämmen. Aber was passiert mit den bisherigen Archivdateien? Was passiert mit den vielen Millionen MXF-Dateien, die in den letzten 10 Jahren erzeugt wurden und bereits heute in den Fernseharchiven lagern? Aus der Sicht des Autors sollten MXF-Archivbestände systematisch und für den Bild- und Ton mathematisch verlustfrei in die neuen MXF-Profile überführt werden. Die Dateien liegen dann in dem Zustand vor, in dem sie schon von Anfang an hätten sein sollen.

“Anwender sollten nicht erwarten, dass zukünftige Produktgenerationen noch mit den ‚Kinderkrankheiten‘ der frühen MXF-Generationen umgehen können.“

Es bedarf einer hochperformanten Konvertierungslösung, die bisherige MXF-Formatvarianten ohne Videoqualitätsverlust in die neuen

MXF-Profile überführt. Über die letzten Jahre haben die Ingenieure bei Cube-Tec tausende von MXF Files von Hand analysiert und Reparatur-Regeln für die unterschiedlichen Fehlerkonstellationen entwickelt.

MXF Legalizer liefert ARD-ZDF MXF-Profil konforme Dateien. Dies ist nun durch eine IRT-Zertifizierung belegt. Die Zertifizierung ist ein wichtiger Schritt zur Qualitätssicherung.

Erkenntnisse aus dem DAVID-Projekt für neu zu entwickelnde Medienstandards

Eine Teilaufgabe im DAVID-Projekt war Konzepte zu entwickeln, wie ein mögliches wenig fehleranfälliges Mediendateiformat aussehen könnte, wenn es auf der grünen Wiese neuentwickelt werden würde. Erfreulicherweise wurden einige dieser Ideen bereits in einem neuen Format umgesetzt. Das größte Standardisierungsprojekt im Bereich neue Mediendateiformate ist bei der SMPTE das Interoperable Master Format (IMF). Die IMF Standardisierung wurde in 2009 von den führenden Hollywood-Studios angeregt um ein einheitliches Austauschformat für Filmmaster-Dateien zu kreieren. Cube-Tec arbeitet bei der SMPTE bereits seit 5 Jahren an der IMF-Standardisierung mit.

Aus einem hochauflösenden (Mezzanine) Videodateiformat sollten alle Distributionsformate (Blu-ray Disc, DVD, on-Air Version, Director's Cut...) erzeugbar werden, so dass alle nationalen Produktvarianten mit ihren vielfältigen Sprach- & Untertitel-Versionen leicht daraus abgeleitet werden können.

IMF setzt auf dem MXF-Format auf. Es vereinfacht jedoch die Formatstruktur, in dem es Audio- und Videostreams trennt und alle Zusammenhänge mittels Metadaten beschreibt, die außerhalb der eigenen MXF in verschiedenen XML-Dateien abgelegt sind (Compo-

sition Play List (CPL), Output Profile List (OPL), Packing List, Asset Map,...). Für den Transport werden dann die einzelnen Dateien wie in einem ZIP-Packer wieder in eine Datei (das Interoperable Master Package - IMP) zusammengefügt.

Warum sollte das Interoperable Master Format nun für Rundfunk- und große AV-Medienarchive zukünftig interessant sein?

Weltweit gibt wahrscheinlich maximal einhundert "MXF-Experten", aber es gibt viele Millionen "XML-Experten" und unendlich viele XML-Tools. Formatprobleme werden durch externe XML-Dateien leichter nachvollziehbar, als wenn die Struktur (in MXF-typische KLV-Elemente) zerlegt werden muss. Es gibt weniger Freiheitsgrade, also weniger Möglichkeiten das gleiche auf verschiedenen Wegen zu erreichen und die Standarddokumente sind leichter verständlich (zumindest ein klein wenig). Das eigentlich spannende ist, dass wir nun in einer EBU-Arbeitsgruppe an einer Version ("App") für den Einsatz bei Rundfunkanstalten arbeiten. Bei diesen Arbeiten werden wir den Fokus „Archiv“ sicherlich nicht aus den Augen verlieren. Damit Anwender zukünftig nicht erst wieder 10 Jahre problematische Mediendateien archivieren, hat Cube-Tec frühzeitig ein Korrekturwerkzeug für IMF auf der NAB Show 2016 in Las Vegas vorgestellt um auch fehlerhafte IMF-Packages automatisiert korrigieren zu können.

Zusammenfassung

Bei Einhaltung anerkannter Archivierungsstandards (zwei räumlich getrennte Online-Kopien und eine externe Offline-Kopie) sind aus heutiger Sicht kaum Datenverluste durch defekte Datenträger oder defekte Datenmigrationen zu erwarten.

Auch wenn es gelingt Mediendateien fehlerfrei zu bewahren heißt das noch lange nicht, dass sie auch zukünftig zur Nutzung zur Verfügung stehen.

Mediendateien die während der Archivierung noch problemlos nutzbar waren, werden obwohl sie bit-identisch aufbewahrt wurden trotzdem häufig nicht mehr lesbar sein.

Dieser Sachverhalt ist für viele Medienarchive heute noch nicht erwartungskonform!

Wenn Mediendateien nicht problemlos maschinenfind- und -lesbar sind, muss der Content schon als sehr wertvoll eingestuft werden, sonst werden diese Mediendateien zukünftig als praktisch nicht existent eingestuft werden.

Auch dieser Sachverhalt ist für viele Medienarchive heute noch nicht erwartungskonform!