

El projecte PREFORMA. Funcionalitats i potencialitats en la perspectiva del Centre de Recerca i Difusió de la Imatge (CRDI)

*David Iglésias Franch, Sònia Oliveras Artau, Pau Saavedra Bendito
Centre de Recerca i Difusió de la Imatge (CRDI)*

El projecte [PREFORMA](#) (PREservation FORMAts for culture information/e-archives) és un projecte de Contractació Pre-Comercial cofinançat per la Comissió Europea dins del Programa FP7-ICT que es va duu a terme des de 2014 fins a 2017. L'objectiu principal del projecte era proveir a les institucions patrimonials que preserven documents digitals de les eines necessàries per a un control total en el procés de recepció d'aquests documents. El resultat obtingut és un conjunt d'aplicacions de codi obert per tal de comprovar formats de fitxers de qualitat amb les següents funcions principals:

- Verificar el compliment amb les especificacions estàndard del format.
- Verificar els criteris d'acceptació de les institucions.
- Crear informes dels fitxers analitzats.
- Realitzar correccions simples.

El tipus de projecte implicava l'agrupació de diferents institucions per a realitzar una contractació amb un objectiu comú, el desenvolupament d'aquestes aplicacions. Amb aquesta finalitat es va voler crear un consorci que inclogués institucions europees de l'àmbit patrimonial juntament amb institucions i empreses de l'àmbit tecnològic.

El consorci estava format per quinze socis de nou països. D'aquests socis, nou eren institucions culturals (institucions patrimonials, biblioteques, arxius, centres d'audiovisuals) tant d'àmbit nacional com d'àmbit local. El consorci també comptava amb tres centres de recerca i tres consultores especialistes en projectes europeus encarregades de donar suport a la licitació, de la coordinació tècnica, de la comunicació i de la sostenibilitat del projecte. La coordinació tècnica era compartida entre una empresa consultora i l'Arxiu Nacional de Suècia (Riksarkivet).

El Servei de Gestió Documental, Arxius i Publicacions (SGDAP) de l'Ajuntament de Girona, com a institució patrimonial en aquest projecte, va participar en les següents tasques: definició dels requeriments tècnics de les aplicacions per a la licitació, valoració de les propostes presentades a la licitació, avaluació dels productes resultants i contribució al banc de proves de fitxers.

El projecte es va estructurar en les tres fases previstes pel desenvolupament de les aplicacions per part de les empreses que s'havien de contractar: disseny, prototipatge i proves. No obstant, el treball dels diferents socis va començar amb la preparació dels requeriments tècnics per a la licitació. Aquí van intervenir les diferents institucions patrimonials agrupades segons els tres àmbits constituïts a partir del tipus de format de fitxer: text, imatge fixa i audiovisuals, juntament amb els especialistes de programari lliure.

L'elecció dels formats de fitxer a verificar i en base als quals es desenvoluparien les aplicacions, vindria feta per part dels socis del projecte, tot i que el desenvolupament havia de permetre poder anar afegint formats, segons exigències de mercat o necessitats de certes institucions. Inicialment es van definir les característiques comunes que tindrien els formats escollits. Així doncs els formats havien de complir amb els següents punts:

- Sense compressió o amb compressió sense pèrdues.
- Lliure.
- Ben documentat.
- Àmpliament utilitzat.

Després de fer un estudi de mercat (Figura 1) i complint amb els punts anteriors es va decidir escollir el PDF/A per a fitxers de text, el TIFF per a fitxers d'imatge fixa i pel que fa als fitxers d'audiovisuals, no es va trobar cap format que clarament complís amb els punts anteriors. En aquest cas, es va deixar que les empreses que presentaven les propostes a la licitació ho plantegessin. El format d'audiovisuals té una complexitat afegida respecte als altres tipus de formats, perquè en realitat es tracta de la combinació de tres codificacions: una pel so, una per les imatges i una pel contenidor d'ambdós.

	AUDIO VISUAL			TEXT	IMAGE
	broadcast	film	Consumer		
PREFORMA stakeholders	MPEG-IMX (MXF/MPEG2) XDCAM HD422 (MXF/MPEG4)	DPX DCP (MXF/JPEG2000)	MOV/MPEG2 AVI/MPEG2 MPEG/MPEG2 MPEG/MPEG4 -AVC	PDF 1.4 PDF/A1	TIFF 6.0 JPEG JPEG2000 RAW
Industry standards	AS107 (MXF/MPEG2) (MXF/JPEG2000) FIMS (MXF/MPEG2)	DCDM (TIFF 6.0) DCP (MXF/JPEG2000) IMF (MXF/MPEG4)	MPEG-AF	PDF	JPEG2000 TIFF
Open standards	MKV/FFV1 OGG/Dirac	PNG	WebM/VP8 OGG/Theora	PDF/A1 PDF/A3 PDF/A3	PNG
PREFORMA	MKV (?) OGG JPEG2000 (?) FFV1 Dirac LPCM (?)			PDF/A	TIFF 6.0 (?)

Figura 1 Taula resum dels usos de formats. Realitzada per Bert Lemmens (2014).

Un cop es van haver escollit els formats es va continuar amb l'elaboració dels requeriments de les aplicacions a contractar. Els mateixos socis que van participar en el procés de preparació dels requeriments, basats en un sistema de puntuació, van avaluar les diferents propostes rebudes. Es va utilitzar aquest sistema d'avaluació per escollir dues propostes guanyadores per a cada àmbit de format. Es van seleccionar sis empreses. Aquestes empreses van realitzar la fase de disseny de les aplicacions desitjades pel tipus de format que havien presentat a la proposta. Aquest disseny

inicial es va tornar a avaluar per part dels mateixos revisors i d'aquesta segona revisió es va seleccionar el millor disseny en cada àmbit de format. Els seleccionats van passar a les següents fases de prototipatge i proves. Les empreses semifinalistes i finalistes de la primera i segona avaluació van ser les següents, segons les classes de documents:

- Text (format PDF/A)
 - veraPDF Consortium, amb el producte final [veraPDF](#)
 - [Preservica](#)
- Imatge fixa (format TIFF)
 - [Easy Innova](#), amb el producte final [DPF Manager](#)
 - [LIBIS](#)
- Audiovisuals
 - [MediaArea](#), amb el producte final [MediaConch](#) - CONformance Checking, amb la proposta de formats MKV|FFV1|LPCM
 - UCL, amb la proposta de formats MKV|JPEG2000|LPCM

En la fase de prototipatge les empreses van desenvolupar l'eina per a verificar el format de fitxer corresponent i a la vegada una interfície per comunicar-se amb les altres dues solucions dels dos formats restants. D'aquesta forma qualsevol de les tres eines és capaç de verificar els tres formats de fitxers, usant els mòduls corresponents.

Durant la fase de prototipatge es va anar fent el seguiment del desenvolupament per part del socis del consorci. I ja un cop les aplicacions van ser acabades, amb el banc de fitxers de proves es va passar a avaluar-les en la darrera fase. L'objectiu d'aquesta fase era identificar possibles errors o millores per acabar de polir les eines.

L'Ajuntament de Girona es va veure molt implicat des del primer moment en la participació d'aquest projecte perquè viu i entén la problemàtica que atacava el projecte i disposa de molts fitxers dels àmbits que afrontava. Des dels fitxers de text ja sigui de documentació històrica, com els d'actual creació, fins als d'imatge fixa o d'audiovisuals del Centre de Recerca i Difusió de la Imatge (CRDI). Es disposa d'un ampli ventall de fitxers per col·laborar en el banc de proves i també es comptava amb una àmplia experiència amb aquests formats. Pel que fa al format de fitxer escollit finalment per l'àmbit d'audiovisual cal dir que no s'utilitza en el CRDI. No obstant, es va poder seguir col·laborant. És remarcable la col·laboració amb Easy Innova, spin off de la Universitat de Girona. Aquesta empresa es va presentar en la licitació i va resultar l'empresa amb millor disseny per implementar el verificador dels fitxers TIFF, per a la imatge fixa. La col·laboració es va produir principalment en dos aspectes concrets. Primer, en el moment inicial, quan Easy Innova realitzava un estudi sobre el flux de treball d'un arxiu per tal de poder presentar la proposta. Posteriorment, durant el desenvolupament de l'aplicació hi va haver una comunicació continuada per tal de valorar els resultats assolits. Es van atendre requeriments concrets, més vinculats al coneixement de la gestió de la imatge d'arxiu, i es van plantejar diferents qüestions referents a formats de presentació de paràmetres i resultats. Una col·laboració enriquidora per a les dues parts i que expliquem de manera més detallada en aquest text.

DPF Manager

El desenvolupament d'un programa com el DPF Manager per a la imatge fixa, i més concretament el TIFF, suposa una fita de gran rellevància en l'escenari de la preservació digital de la fotografia. Suposa, al mateix temps, un primer pas en la direcció desitjada per als arxius i aquelles institucions responsables del patrimoni en imatge. Cal tenir en compte que la realitat dels arxius està formada també per altres formats gràfics, com són el JPEG2000, els formats RAW i tota l'extensa família JPEG.

Per posar les coses en context, és important situar adequadament el rol del TIFF en un escenari concret, en aquest cas un arxiu fotogràfic com el del CRDI. Això ens ha de permetre valorar la importància del programa en qüestió i, també, qüestionar en certa manera l'ortodòxia del TIFF, sense discutir-ne l'estatus adquirit d'estàndard de facto per a la preservació digital de la imatge fixa.

En el CRDI la presència de fitxers TIFF és notable però el percentatge de representació respecte a altres formats és decreixent. Aquesta circumstància s'explica sobretot pels fluxos de treball generats a partir dels canvis en la digitalització i, sobretot, a partir dels acords amb un diari d'àmbit nacional que impliquen la gestió d'un arxiu digital de premsa. A grosso modo, podem explicar que amb l'inici de la digitalització massiva de fotografia, a principi d'aquest segle, es van crear uns màsters digitals en TIFF, sense compressió. Aquestes imatge es creaven a partir de la digitalització amb escàners de diferents tipus i es transformaven a JPEG per a les còpies d'ús quotidià. També les imatges que provenien de reportatges d'encàrrec s'entregaven en aquest format. Aquesta pràctica es va mantenir durant més de deu anys, arribant a una xifra aproximada de tres-cents mil fitxers. El canvi de plantejament es va produir a partir de la digitalització amb càmera i la possibilitat de fer captures en RAW. El flux de treball en la digitalització passava, a partir d'aquest moment, per una captura inicial amb format RAW, editat posteriorment i transformat a JPEG com a imatge d'ús final, tant per a la catalogació com per a la publicació on-line. Començava així una nova etapa per a l'arxiu, en què el tàndem RAW-JPEG s'imposava al TIFF, bàsicament per a l'acompliment de les funcionalitats principals assignades i, més concretament, pel fet de comptar amb fitxers reversibles per a l'edició (els RAW de càmera), pel fet de comptar amb fitxers que integren contenidors XMP per a les metadades (JPEG), i pel fet de comptar amb fitxers compatibles amb qualsevol navegador i per tant que poden ser publicats als catàlegs web. No entrarem a explicar ara tot l'argumentari que ens va portar a aquesta decisió. En canvi, sí que hem de fer esment d'un fet molt significatiu: el fet d'abandonar els TIFF ens obria una escletxa en l'estabilitat de l'arxiu digital.

L'altra circumstància que va portar a la pèrdua del domini del TIFF va ser l'acord amb el diari El Punt-Avui per gestionar el seu arxiu digital. Aquest acord es va concretar, en un primer moment, en la transferència de prop de 2 milions d'imatges, totes elles en format JPEG, i el compromís de transferir mensualment la producció diària, que suposa una ingesta anual d'aproximadament dues-centes cinquanta mil imatges, totes elles JPEG, un cop feta la selecció. Per tant, una vegada més, s'imposa la lògica del mercat, la producció dels JPEG en un entorn al marge de les institucions patrimonials.

En definitiva, aquestes circumstàncies, que no són exclusives del CRDI, expliquen bé el sentir general del desajustament entre la producció externa dominat al mercat i les

necessitats en matèria de preservació, en allò referent als formats. No entrarem ara a valorar totes aquestes qüestions, però sí que volem alertar de la necessitat de no combregar amb els postulats més assentats sense una reflexió prèvia, ja que cal fixar els criteris segons el context en què es trobi cadascú. Sí que són exclusives del CRDI, però, les decisions preses entorn al flux de treball i, en aquest sentit, val a dir que responen més a un pragmatisme profundament argumentat que a una teorització sense esclatxes. Perquè les esclatxes hi són, i giren sobretot a l'entorn de custodiar formats RAW propietaris i a la limitació de característiques tècniques dels JPEG.

En qualsevol cas, aquesta introducció per explicar el context de creació i custòdia d'imatge digital per part del CRDI ha de servir per valorar aquest nou software en la seva justa mesura. Ja dèiem que el seu desenvolupament marcava un camí en la preservació dels formats gràfics. Ara podem afegir que la seva implementació pot tenir uns efectes molt beneficiosos encara que parcials, en el sentit que la creació de més del 99% d'imatges que es produeixen arreu són en format JPEG. Aquest fet té lògicament un reflex sobre el patrimoni. Establint un paral·lelisme amb la fotografia química, ens podríem preguntar si algú podria plantejar-se la custòdia d'arxius fotogràfics del segle XIX sense tenir en compte les còpies a l'albúmina. I encara més, podríem al mateix temps centrar-nos en aquests temps remots ignorant l'existència del col·lodió humit? Som conscients que els paral·lelismes, en aquest sentit, són complicats, i que requereixen d'un cert atreviment. No obstant, reflecteixen bé la idea principal, la del testimoni de la tecnologia d'una època concreta. En els nostres temps, els RAW i els JPEG estan marcant també una època.

Fetes aquestes consideracions, podem centrar-nos ara en el format TIFF, el subjecte entorn al qual gira el desenvolupament del programa. Sabem que el format va ser creat per la companyia Aldus Corporation el 1986, en un intent de superar la inexistència d'un estàndard per a la imatge. El 1987 va passar a mans d'Adobe que va elaborar noves versions: la 4.0 el 1987, la 5.0 el 1988, i la 6.0 el 1992. En aquests anys el format es va consolidar com a estàndard de facto i ha anat adquirint noves funcionalitats, en el que són les extensions. No obstant, el seu gran valor resideix en la capacitat de descriure imatges de color sense comprimir, en la seva adaptabilitat als canvis, en l'ús extens i prolongat del format, en la compatibilitat amb software, etc., fet que el converteix en un format òptim per a l'arxiu.

Les variants dins el mateix format són sovint problemàtiques per a la preservació ja que les extensions no són sempre coincidents i, a més, els desenvolupadors de software interpreten de manera diferent les especificacions tècniques. Per això, és destacable l'esforç que es va fer per definir un TIFF de preservació, anomenat TI/A, una tasca estretament vinculada a aquest projecte. El TI/A (Tagged Image for Archival)ⁱ és una iniciativa de la Universitat de Basilea que pretén fixar uns recomanacions en favor de la preservació del format. Ja va circular un primer esborrany per al comitè ISO, tot i que l'estàndard està encara pendent d'aprovació. El DPF Manager es va desenvolupar prenent el TI/A com a referència a l'hora de validar qualsevol fitxer TIFF. En aquest sentit, és una iniciativa única i pionera per a les aplicacions amb capacitat d'identificar i validar formats.

Centrant-nos en el DPF Managerⁱⁱ, el podem descriure com una aplicació multiplataforma creada per a la validació i l'anàlisi d'imatges TIFF amb l'objectiu de

tenir un control important sobre aquest tipus de fitxer en el moment de la ingesta. L'aplicació permet visualitzar l'estructura del fitxer i mostrar el conjunt de metadades que continguin, ja siguin tècniques o documentals. A més, permet fixar criteris d'acceptació en base a paràmetres tècnics de la imatge, la qual cosa facilita el control qualitatiu durant la ingesta. El programa compte també amb la possibilitat de correcció de determinat errors que freqüenten en molts d'aquestes fitxers. Entre aquests errors trobem, per exemple, text que no està codificat en ascii de 7 bits, que és la codificació permesa pels TIFF. També corregeix la inconsistència de metadades de diferents contenidors, com poden ser els XMP, EXIF o IPTC. Tota la informació que s'adquireix en l'anàlisi es presenta també en un informe, fet que permet disposar d'una lectura més amable del conjunt de la informació obtinguda.

Al marge del coneixement general del programa i de les seves funcionalitats, explicades de manera genèrica en el paràgraf anterior, ens sembla interessant incloure en aquest text les peticions i suggeriments que es van fer des del CRDI durant el desenvolupament del programa i les reflexions fetes sobre aspectes concrets que poden ser d'interès general pel sector de la imatge.

Una de les reflexions més interessants que es van fer van ser al voltant de com es presentaven les estructures dels fitxers analitzats, sobretot tenint en compte l'experiència anterior amb altres programes que tenen una orientació similar, com poden ser el Jhove i el DROID i que, per tant, permetien una comparativa sobre la informació que ofereixen. En el cas de Jhove, per exemple, s'observa que la presentació de l'estructura del fitxer és diferent. La divergència pot estar en la identificació de directoris. Si ens centrem en el contenidor metadades EXIF, observem com els dos programes el presenten a diferent nivell de l'estructura. De fet, el DPF Manager mostra l'estructura real del fitxer, segons els directoris (IFD). El Jhove, en canvi, mostra l'EXIF com si fos un IFD independent, però en realitat es troba dins un IFD i, per tant, es considera un sub-IFD. També a nivell d'estructura i referent als contenidors de metadades va ser interessant aprendre com els XMP, IPTC, EXIF i Perfils ICC segueixen l'estructura dels estàndards respectius i que contenen tots una referència (offset) que els permet localitzar dins els TIFF. En aquest aspecte podem dir que el DPF Manager presenta una estructura més ajustada a la realitat dels fitxers i que en permet una comprensió més clara.

Entre les peticions que es van fer, hi ha la de mostrar la informació del perfil de color (ICC Profile), per ser una dada rellevant per a l'arxiu. És ben conegut que la diversitat en perfils de color en la captura i la sortida fa recomanable la utilització de perfils abstractes que no estiguin vinculats a cap dispositiu i que, al mateix temps, puguin representar un ventall ampli de colors. En aquest sentit l'Adobe RGB (1998) ha estat utilitzat abastament. Per això calia identificar els perfils, quelcom que es va incloure en la versió final.

Una altra de les peticions estava relacionada amb la presentació dels valors continguts en les metadades, per tal de poder identificar correctament de quina metadada estem parlant. Per exemple, en el cas d'una imatge d'un arxiu francès, podem trobar aquesta informació en un dels camps "Paul Poiret (1879-1944), couturier et decorateur français". Només podem saber que es refereix al títol si s'identifica la metadada corresponent. Si fem la cerca en la metadada del fitxer, obtenim el següent:

photoshop:Headline>Paul Poiret (1879-1944), couturier et decorateur français
</photoshop:Headline>. Per tant, d'alguna manera s'havia d'indicar la metadada de pertinença, ja que no sempre és evident.

LX00103.tif
F:\03 C R D \033-Activitats\Postgrau_Fotografia\Postgrau on-line\Iglesias-David\3-Edicion\Actividad4\LX00103.tif
⚠ Baseline TIFF 6.0

	Errors	Warnings
Baseline TIFF 6.0	3	1

IPTC

Name	Value
By-line Title	
Category	
Keywords	Teatros
Country/Primary Location Code	
Originating Program	FotoWare FotoStation
Coded Character Set	%G
By-line	Foto Lux
Caption/Abstract	Entrada del teatro Alb��niz, en la plaza Sant Agust��. En los carteles del exterior se anuncia la actuaci��n de los payasos Lon Chaney.
City	
Object Name	[Teatro Alb��niz]

File structure

- IFD - Main image
 - EXIF
 - XMP
 - IPTC
 - ICC
 - Description: Phase One Gray G2.2
 - Creator: SampleICC
 - Version: 2.1
 - Class: Display
- IFD - Main image

Conformance checker

⚠ Baseline TIFF 6.0

Type	ID	Location	Description
✖	IDFE-0002	tag 270 ImageDescription	Only 7-bit ASCII codes are accepted
✖	IFDI-0004	IFD2	Image IFD must have tag X Resolution
✖	IFDI-0005	IFD2	Image IFD must have tag Y Resolution
⚠	TAG-284-0005	IFD1	PlanarConfiguration is irrelevant if SamplesPerPixel is 1, and need not be included.

Metadata analysis

Description
✖ Author on TAG and IPTC in IFD 1 (Foto Lux,)
✖ Author on IPTC and XMP in IFD 1 (, Foto Lux)

Figura 2. Visualitzaci   parcial de la informaci   obtinguda en la validaci   d'un fitxer TIFF amb el DPF Manager.

Per entendre millor totes aquestes q  estions i les funcionalitats de l'aplicaci  , expliquem a continuaci   la validaci   del fitxer TIFF de la Figura 2. Per a l'an  lisi d'aquesta imatge s'han escollit les especificacions t  cniques del TIFF Baseline 6.0. Les altres opcions eren el TIFF/IT, TIFF/EP i Extended TIFF 6.0. Si ens fixem amb l'estructura, observem que el fitxer cont   dos directoris (IFD) que fan refer  ncia a dues imatges diferents, la principal i la del thumbnail. A m  s, l'IFD de la imatge principal cont   subdirectoris de metadades i, m  s concretament: EXIF, XMP, IPTC i ICC. Encara que es tracti de contenidors independents, f  sicament es troben dins un IFD. Cadascun d'ells cont   metadades, que podem desplegar i veure'n els valors. En el cas de l'ICC, abans comentat, ens proporciona informaci   del perfil de color. Entre els errors detectats, observem un error com   referent a la codificaci   del text, ja que aquesta varia en les metadades del diferents contenidors (EXIF, IPTC, XMP) i, en qualsevol cas, la codificaci   del text dep  n del software utilitzat. Referent als errors trobats en l'an  lisi de metadades, hi ha un problema de consist  ncia amb el camp autor en els

diferents contenidors IPTC i XMP, un problema originat pel software utilitzat per a la creació de la imatge.

Per últim, comentar un qüestió clau que es va plantejar amb la possibilitat d'incloure el DNG basant-nos en la consideració de que es tracta d'un RAW d'Adobe amb estructura TIFF. Davant la tendència en l'àmbit patrimonial a conservar els RAW com a fitxers de preservació en substitució dels TIFF, semblava apropiat adaptar el programa per a l'anàlisi dels DNG. Malgrat l'acord en aquest punt, el projecte no oferia suficients recursos per al desenvolupament d'aquest validador. Val a dir però, que canviant l'extensió .dng a .tiff el validador llegeix els tags comuns d'aquests formats, ja que l'estàndard TIFF/EP es basa en el DNG, és a dir, recull bona part de la informació de captura. No obstant, cal remarcar que no es pot pensar en una validació correcta del DNG ja que no es pot obtenir informació dels tags propis d'aquest format RAW. La validació de fitxers RAW queda pendent doncs per a futurs projectes. En qualsevol cas, podem dir que amb l'assoliment del DPF Manager, el camí està traçat.

Mediaconch

Pel que fa al vídeo digital, no existeix un ampli consens sobre quin és el millor format que garanteixi la preservació dels documents audiovisuals a llarg termini, per això, abans d'implementar el programa, s'havien estudiat les propietats dels diferents formats. Cal dir que en el cas del vídeo digital, hem de contemplar els tres elements que conformen el fitxer, això és, el còdec de vídeo, el còdec d'àudio i el format contenidor. L'elecció final va ser el còdec de vídeo FFV1 (FF Video Codec 1), en la seva versió 3, amb el còdec d'àudio LPCM (Linear Pulse Code Modulation), encapsulats en un format contenidor Matroska.

Un dels requeriments inicials era que tant el format contenidor com els còdecs havien de ser de codi lliure i no propietaris, la qual cosa ja descartava moltes altres opcions. Un altre important requeriment era que els còdecs fossin sense pèrdues. En el vídeo digital es diferencien els còdecs sense pèrdues (*lossless*) dels còdecs amb pèrdues (això també és aplicable a l'àudio). Els primers es basen en un tipus de compressió que no sacrifica part de la informació en el procés de codificació per tal de reduir el pes dels fitxers. Tenen l'inconvenient que no poden arribar a tant altes taxes de compressió com els còdecs amb pèrdues, però asseguren que la informació audiovisual del fitxer romandrà inalterada en el futur. La majoria dels fitxers de vídeo digital que formen part dels arxius hauran de ser migrats en algun moment del futur a causa de l'obsolescència dels formats, per això, és convenient assegurar-se que en aquestes futures migracions no es produiran pèrdues d'informació que afectin a la qualitat o integritat dels fitxers.

Un dels problemes de l'ús de còdecs *lossless* per a les institucions, sobretot per a aquelles de mida mitjana o petita, és el fet d'haver de treballar amb fitxers molt pesats que ocupen molt d'espai d'emmagatzematge i requereixen una gran potència de processament. Veiem alguns exemples:

- Uncompressed V210 10-bit: 100 GB per hora de vídeo aprox.
- Lossless FFV1 o JPEG 2000 a 10 bit: 45-50 GB per hora de vídeo aprox.

- MPEG-2 (50 Mbps): 25 GB per hora de vídeo
- DV (25 Mbps): 12 GB per hora de vídeo
- MPEG-2 (qualitat DVD): 3,6 GB per hora de vídeo aprox.

Ara bé, si tenim en compte que la capacitat d'emmagatzematge dels dispositius no ha parat d'augmentar i els costos han disminuït, hauríem de ser capaços de superar aquest problema.

Pel que fa al nostre arxiu, el còdec *lossless* FFV1 no hi és present. En part, degut als costos d'emmagatzematge que ja em esmentat i, en part, perquè es tracta d'un còdec encara bastant recent. L'arxiu audiovisual del CRDI, pel que fa als fitxers de vídeo digital, està format principalment pels següents formats: AVI/DV i MPEG-2 per al vídeo digitalitzat d'origen analògic, i MPEG-4/AVC (H.264) per al vídeo nascut digital, ja que aquest és el format en què ingressa a l'arxiu el vídeo generat per les televisions amb les quals tenim conveni.

Una vegada feta l'elecció, l'objectiu va ser desenvolupar una aplicació de software de codi lliure capaç de testejar i validar els fitxers de vídeo en aquest format per a la seva preservació a llarg termini. El resultat es va concretar en el programa MediaConch, desenvolupat per l'equip de MediaArea, que també és responsable de la creació del conegut programa de codi obert MediaInfo per a la comprovació de fitxers multimèdia. Tot el software desenvolupat per MediaArea durant el projecte PREFORMA s'ofereix amb llicències de codi obert com la GNU General Public License 3.0.

A més, l'equip de MediaArea està dedicat al desenvolupament de la normalització dels formats Matroska i FFV1 per garantir la seva longevitat com a format de preservació digital recomanat.

La discussió sobre els formats

Com hem dit, hi ha diversos còdecs *lossless* al mercat, però si ens quedem només amb els de codi lliure les opcions es redueixen. Després de les primeres valoracions, van quedar tres principals combinacions de còdecs i formats contenidors:

- MXF/JPEG2000/LPCM: un format relacionat amb la indústria que tracta la preservació d'emissió de televisió i material cinematogràfic.
- MKV/FFV1/LPCM: un format de codi obert destinat a la preservació de material audiovisual.
- OGG/Dirac/LPCM: un altre format de codi obert destinat a la preservació de material audiovisual.

En el simposi “*No time to wait! Standardizing FFV1& Matroska for Preservation*”ⁱⁱⁱ que va tenir lloc el juliol de 2016 a Berlín, es van posar sobre la taula els arguments pels quals es va acabar triant la combinació Matroska/FFV1/LPCM. El simposi va reunir arxivers audiovisuals i dissenyadors/implementadors de formats audiovisuals, amb l'objectiu de trobar un format estàndard per a la preservació d'aquest tipus de documents. En aquest simposi va tenir especial rellevància el treball del CELLAR (Codec Encoding for LossLess Archiving and Realtime transmission), un grup de treball l'objectiu del qual és trobar un format estàndard transparent, autodescriptiu i

sense pèrdua, capaç de preservar els documents audiovisuals de l'obsolescència tecnològica.

Entre els arguments a favor de l'elecció de Matroska, es valora el fet que és un format contenidor d'audiovisuals de llicència oberta amb funcions àmplies i flexibles, i una comunitat d'usuaris molt activa. Les seves especificacions estan publicades i es poden consultar lliurement. Des de la seva creació el 2002, el format Matroska ja incorporava l'objectiu d'esdevenir un estàndard. Pot contenir una gran varietat de còdecs de vídeo i d'àudio, així com metadades tècniques i descriptives, i integra una estructura de metadades jeràrquica i flexible, basada en EBML, que és un metallenguatge simplificat basat en XML. Ha estat àmpliament utilitzat a Internet i ha estat adoptat pel projecte WebM de Google com a format de vídeo per a streaming a través d'Internet.

Per la seva part, FFV1 és un còdec de vídeo sense pèrdues, eficaç i ràpid en el procés de codificació, dissenyat de manera que respon als requisits de preservació digital. Potser un dels principals problemes associats al còdec FFV1 és que no hi ha al mercat moltes aplicacions capaces d'interactuar amb ell, tot i que es pot transcodificar amb la llibreria FFmpeg de codi obert.

Pel que fa al flux d'àudio, PCM és una forma de codificar àudio sense compressió que consisteix en prendre mostres del senyal analògic a intervals regulars. LPCM és un tipus específic de PCM on els nivells de quantificació són uniformes.

L'adopció del còdec FFV1 en institucions patrimonials no ha parat de créixer, sobretot a partir del desenvolupament de la versió 3 entre 2010 i 2015 ^{iv}. És el cas, per exemple, de la Mediateka Austríaca ^v que fou una de les primeres institucions en adoptar el còdec de vídeo FFV1, en aquest cas en un contenidor AVI (Audio Video Interleaved), per al material capturat d'una font analògica. Argumenten a favor del còdec *lossless* FFV1 que és fàcilment accessible gràcies al projecte FFmpeg i que és més ràpid i simple que el còdec *lossless* JPEG2000. En aquest sentit, la Mediateka Austríaca ha fet comparacions amb altres còdecs com JPEG2000, Dirac o *lossless* H.264 i n'ha resumit els pros i contres ^{vi}.

Pel que fa a JPEG2000, el consideren un còdec teòricament viable per a la conservació de documents de vídeo digital a llarg termini, ja que compleix els requisits de ser un còdec *lossless* estandarditzat, però exposen com a inconvenients el fet que hi ha poques institucions que conservin els seus arxius audiovisuals en aquest format, que és més lent en la codificació que FFV1, que no existeixen moltes aplicacions que interactuïn amb aquest format i que les seves especificacions són complexes i no són lliurement accessibles.

El còdec Dirac també compleix els requisits plantejats, ja que s'ha desenvolupat amb llicències de codi lliure com GNU i permet la codificació sense pèrdues. De fet, el projecte va ser iniciat per la BBC l'any 2005 amb el propòsit de ser utilitzat en tots els àmbits, tant en la producció, difusió i arxiu de documents audiovisuals. Però també n'assenyalen com a inconvenients la seva complexitat i la seva poca popularitat.

Un altre format que s'avalua és el H.264 en la seva versió *lossless*. És equivalent al còdec conegut com AVC (Advanced Video Codec) que forma part de l'estàndard MPEG-4 i és un codec àmpliament utilitzat en la producció de vídeo digital, gràcies sobretot al fet que ha estat implementat en un gran nombre de càmeres, i que s'utilitza profusament en la difusió de vídeo a través d'Internet. Té una versió *lossless*, tot i que

s'utilitza molt menys que la versió *lossy* (amb pèrdues). Algunes aplicacions que suporten H.264 no suporten la seva versió *lossless*.

També s'ha fet un estudi similar en el marc de les Federal Agencies Digitization Guidelines Initiative (FADGI), plasmat en els documents agrupats sota el títol *Digital File Formats for Videotape Reformatting*^{vii}. En aquests documents s'avaluen els formats contenidors AVI, MOV, Matroska, MXF i MPEG-2, i els còdecs Uncompressed 4:2:2, 8-bit (UYVY and YUY2), Uncompressed 4:2:2, 10-bit (v210), JPEG2000 - Lossless, FFV1 i MPEG-2 - 4:2:2 Profile/Main Level.

Del format contenidor Matroska en destaquen la seva transparència, la bona documentació, la possibilitat d'afegir metadades i la seva independència de patents i drets de propietat intel·lectual. Pel que fa al còdec FFV1, es considera només acceptable en la majoria d'indicadors respecte a altres còdecs com MPEG-2, però se'n destaca el fet de ser de codi lliure i no estar subjecte a drets de propietat intel·lectual.

MediaConch

Com hem dit, MediaConch és el programa que ha resultat del projecte PREFORMA en la seva aplicació als documents audiovisuals. Es pot utilitzar amb línia de comandaments, amb una interfície gràfica d'usuari o amb una interfície basada en web, i pot ser molt útil per a un arxiu audiovisual un cop s'han definit els paràmetres i requeriments que han de tenir els fitxers digitals de vídeo i àudio per a la seva conservació a llarg termini.

Ha estat desenvolupat per l'equip de la companyia MediaArea dedicada al desenvolupament de software lliure i de codi obert, com ho és el programa MediaInfo, un programa àmpliament utilitzat per la comunitat d'arxivers audiovisuals, ja que permet identificar les metadades tècniques de pràcticament qualsevol format de vídeo digital.

L'equip de MediaArea ha volgut comptar, a més, amb la participació de la comunitat d'usuaris com un factor clau per avaluar l'èxit del projecte. Per això, durant la fase de prototip, MediaArea ha realitzat el treball de desenvolupament permetent l'accés públic a través de Github.com.

La part central del programa és el *checker* (testejador) on es determinen a priori els paràmetres que ha de tenir un fitxer model, a partir del qual es poden comparar tots els fitxers del repositori. Es pot avaluar un fitxer local, un fitxer online o tots els fitxers continguts en una carpeta. Els resultats d'aquests tests són presentats en un informe que permet avaluar si els fitxers de vídeo del nostre repositori contenen les característiques requerides per a la preservació a llarg termini o, per contra, presenten errors en la seva codificació.

El *checker* respon a preguntes de forma afirmativa o negativa. Només si totes les respostes són superades es pot considerar que el fitxer és vàlid per a la seva conservació a l'arxiu. L'usuari pot definir quina és la política o perfil que vol aplicar als fitxers de vídeo de la seva institució. Així mateix, un cop definida la política, aquesta es pot compartir amb altres institucions.

A part del perfil principal (MKV/FFV1/LPCM), existeix un llistat de polítiques o perfils predefinits, com per exemple MPEG-4 AVC (H.264-AAC), Memoriv Video Files

Recommendations, CAVPP Preservation Master, etc., però també es poden importar, editar i crear noves polítiques. Existeix a l'aplicació una secció de polítiques on la comunitat d'usuaris pot afegir la política de la seva institució, al mateix temps que es poden importar al llistat.

Aquest és un exemple de les metadades que avalua el *checker* amb la definició de la política de digitalització MKV FFV1.

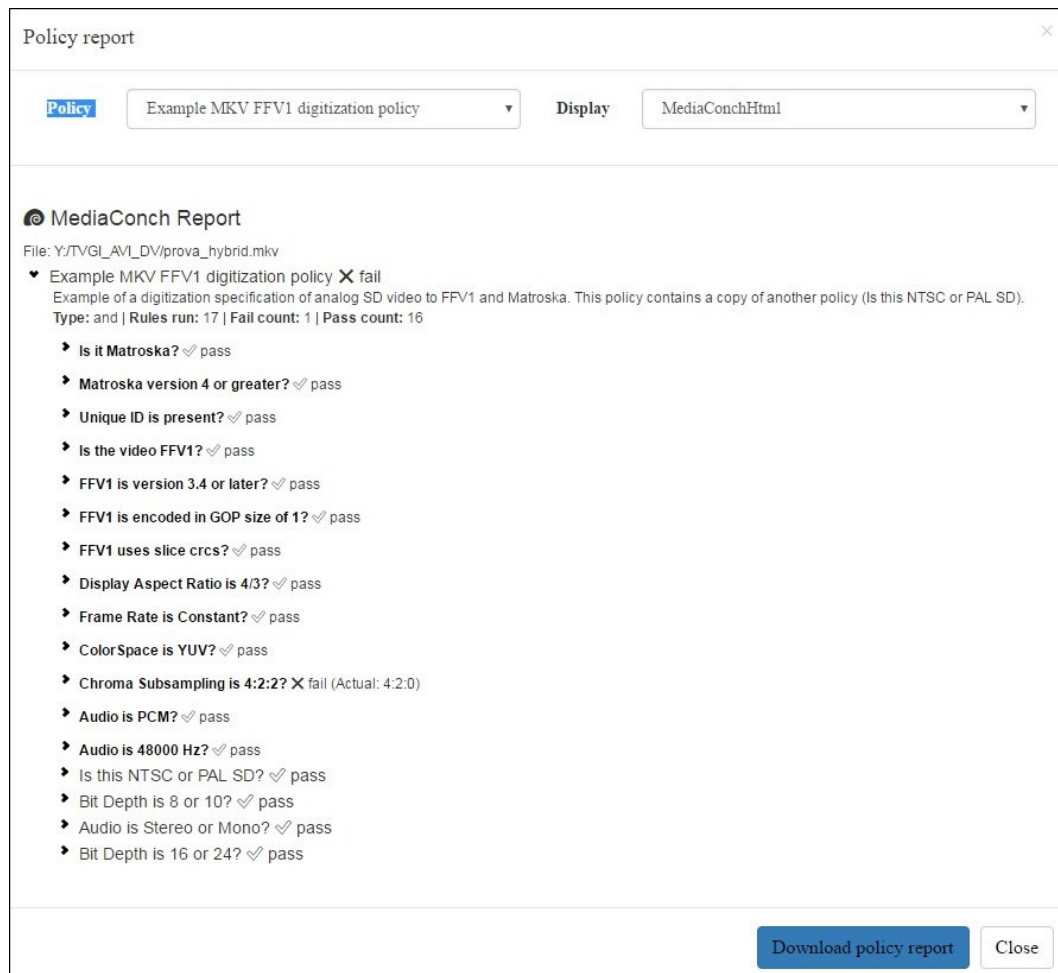


Figura 3. MediaConch report

Els resultats de l'informe es poden presentar en diferents formats (xml, csv, html o txt) i existeix la possibilitat de conservar en una base de dades totes les accions realitzades amb MediaConch.

En definitiva, podem dir que l'equip de MediaArea ha creat un programa d'una gran simplicitat quant a la seva utilització i d'un gran potencial per a desenvolupar tasques de control de qualitat en la digitalització i contribuir així a garantir la pervivència de les col·leccions audiovisuals en format digital.

i Esborrany de l'estàndard (2015): <http://www.digitalmeetsculture.net/wp-content/uploads/2015/04/EASY-INNOVA-4.-TIFF-A-Standard-Proposal.pdf>

ii Descarregar el DPF Manager a: <http://dpfmanager.org/>

iii Media Area. No Time to Wait: Standardizing FFV1 and Matroska for Preservation. <https://mediaarea.net/blog/2016/07/26/No-Time-To-Wait-Preservation-FFV1-Matroska-Symposium>

iv Blewer, Ashley; Rice, Dave. Status of CELLAR: Update from an IETF Working Group for Matroska and FFV1. https://pdfs.semanticscholar.org/8fdc/4cbe2c9ad280add30812bf9496f4fdb86beb.pdf?_ga=2.262748428.1633096483.1538490461-1874169606.1538490461

v The archivist's video codec and container FAQ. Mediateka Austríaca. http://download.das-werkstatt.com/pb/mthk/info/video/FAQ-digital_video_archiving.html

vi Written by Peter B., Hermann Lewetz and Marion Jaks. Comparing video codecs and containers for archives. http://download.das-werkstatt.com/pb/mthk/info/video/comparison_video_codecs_containers.html#codec_tests

vii File Formats for Videotape Reformatting http://www.digitizationguidelines.gov/guidelines/video_reformatting_compare.html