



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

SBA SISTEMA BIBLIOTECARIO
DI ATENEIO

Formati dei file raccomandati per l'archiviazione a lungo termine e per la disseminazione web in Phaidra

A cura di Gianluca Drago

Maggio 2019

Licenza: <https://creativecommons.org/licenses/by-nc-sa/4.0/>

Premessa

Questo documento intende fornire una panoramica sui formati dei file da adottare in funzione di due possibili destinazioni del documento digitale:

- l'archiviazione a lungo termine
- il caricamento in Phaidra e la conseguente disseminazione web

Quando il documento caricato in Phaidra è anche l'unico file salvato, le due destinazioni finiscono per coincidere, ma in generale probabilmente si vorranno produrre due diversi file, in due diversi formati per rispondere alle differenze di requisiti e di utilizzo nelle destinazioni finali.

Nelle tabelle che seguono le raccomandazioni per l'archiviazione a lungo termine sono distinte da quelle relative alla disseminazione in Phaidra.

Non esistono criteri assoluti nella scelta del formato dei file. La scelta è sempre dipendente da differenti valutazioni che chi archivia dovrà fare puntualmente, caso per caso, e risulterà spesso in un compromesso tra la miglior qualità ottenibile e i limiti imposti dai costi di produzione, elaborazione e stoccaggio dei file, nonché, per il pregresso, dall'opportunità di una conversione a nuovo formato.

Questa scelta è particolarmente significativa nell'ottica dell'archiviazione a lungo termine per la quale sono desiderabili una qualità che rispetti l'autenticità e l'integrità del documento originale e un formato che garantisca l'accesso ai dati nel lungo periodo.

Questo documento deve essere visto più come un ausilio alla scelta ragionata che spetta a chi archivia, piuttosto che come un elenco di direttive da seguire alla lettera.

Nelle tabelle che seguono la colonna intestata "Web consigliato" include solo i formati che consentono una visualizzazione diretta nel browser, senza l'ausilio di applicazione esterne. La colonna intestata "Tipo di oggetto in Phaidra" ha la funzione pratica di specificare quale "tipo di oggetto" deve essere selezionato da chi archivia in Phaidra al momento del caricamento.

Per la spiegazione delle abbreviazioni qui utilizzate, si veda la sezione "Abbreviazioni usate" alla fine del documento.

Criteri di scelta del formato

Esistono alcuni criteri generali ai quali ci si può attenere per quanto riguarda la scelta del formato più adatto all'archiviazione. Pur con qualche disomogeneità, archivi e istituzioni che si occupano di conservazione del patrimonio digitale identificano i più importanti in quelli di seguito riassunti.

- **Apertura:** un formato si dice "aperto" quando è conforme a specifiche pubbliche, cioè disponibili a chiunque abbia interesse ad utilizzare quel formato. La disponibilità delle specifiche del formato dovrebbe renderne possibile la decodifica, anche in assenza di prodotti che effettuino tale operazione automaticamente. Alcuni archivi specificano più rigorosamente che il formato non deve essere proprietario¹.
- **Portabilità:** si intende la facilità con cui i formati possono essere usati su piattaforme diverse, sia dal punto di vista hardware che software. Nel definire la portabilità di un formato si tiene in considerazione anche la disponibilità di strumenti che lo rendano accessibile sia in fase di creazione dei file sia quando si deve accedere ai dati. La presenza di dipendenze esterne, di meccanismi di protezione tecnica o di brevetti contrastano con la portabilità.
- **Qualità e funzionalità:** la capacità da parte di un formato di mettere a disposizione funzionalità che assicurino qualità e ricchezza dei dati, buone prestazioni in termini di velocità (ed eventualmente di compressione), inclusione di metadati e dati di differente natura.

¹ Generalmente si intende con formato "proprietario" un formato non aperto, coperto da brevetti o licenze oppure le cui specifiche non siano interamente disponibili.

- **Supporto allo sviluppo:** attiene alle risorse necessarie alla manutenzione e sviluppo del formato nonché dei prodotti informatici che lo gestiscono.
- **Diffusione:** misura il grado di diffusione del formato nel mondo, e in particolare il livello di adozione da parte dei più importanti archivi internazionali. La diffusione di un formato ha implicazioni sulla probabilità che esso venga supportato nel tempo, attraverso la disponibilità di prodotti informatici idonei alla sua gestione e visualizzazione.
- **Trasparenza:** si riferisce al grado in cui la rappresentazione digitale è aperta all'analisi diretta con strumenti di base. La trasparenza è migliore se il contenuto è codificato in codifiche standard. La crittografia è incompatibile con la trasparenza; la compressione inibisce la trasparenza (tuttavia, per motivi pratici, audio e video digitali raramente sono archiviati in una forma non compressa).
- **Auto-documentazione:** gli oggetti digitali che si auto-documentano sono più sostenibili e meno vulnerabili nel lungo periodo rispetto agli oggetti che sono archiviati separatamente dai metadati necessari per renderli utilizzabili. Un oggetto digitale che contiene metadati descrittivi di base (l'analogo del frontespizio di un libro) e incorpora metadati tecnici e amministrativi relativi alla sua creazione e alle prime fasi del suo ciclo di vita sarà più facile da gestire, da monitorare per integrità e usabilità, e da trasferire da un sistema di archiviazione al successivo.

Per approfondimenti sui criteri di scelta del formato si confrontino:

- DPCM 3 dicembre 2013, alleg. 2 “*Regole tecniche in materia di sistema di conservazione*”, ai sensi del “*Codice dell'amministrazione digitale*” https://www.agid.gov.it/sites/default/files/repository_files/leggi_decreti_direttive/dpcm_3-12-2013_conservazione.pdf pag. 17
- LOC, *Sustainability of Digital Formats: Planning for Library of Congress Collections*, 2017 <https://www.loc.gov/preservation/digital/formats/sustain/sustain.shtml#disclosure>
- Evelyn Peters McLellan, *Selecting Formats for Digital Preservation: Lessons Learned during the Archivematica Project*, 2010 https://groups.niso.org/apps/group_public/download.php/4237/IP_McLellan_Selecting_Formats_isqv22no2.pdf
- Universität Wien, *Formats for long-term preservation* <https://datamanagement.univie.ac.at/en/about-phaidra/formats/formats-for-longterm-perservation/>
- NARA, *Frequently asked questions about Digital Audio and Video*, 2016 <https://www.archives.gov/records-mgmt/initiatives/dav-faq.html>
- PACKED, *A short guide to choosing a digital format for video archiving masters*, 2014 <https://www.scart.be/?q=en/content/short-guide-choosing-digital-format-video-archiving-masters>

Testo

Archiviazione Consigliato	Archiviazione Possibile	Web Consigliato	Web Possibile	Tipo di oggetto in Phaidra	Note
PDF/A		PDF/A		Documento (PDF, TeX)	Formato di archiviazione ampiamente utilizzato, che vieta l'uso di alcune funzioni del PDF che potrebbero essere difficili da rendere in futuro. Ben supportato; specifiche interamente disponibili
	PDF	PDF		Documento (PDF, TeX)	In genere, più leggero di PDF/A, ma non adatto alla lunga conservazione
	TeX, LaTeX		TeX, LaTeX	Documento (PDF, TeX)	Formato di marcatura particolarmente usato per la rappresentazione di formule matematiche. Senza i propri file accessori (immagini, indici, bibliografie...) ci può essere perdita di contenuti, formattazione, funzionalità. Per Phaidra è preferibile l'esportazione in PDF/A
	HTML, XHTML		HTML, XHTML	Altro	Senza i propri file accessori (immagini, CSS, JavaScript...) ci può essere perdita di contenuti, formattazione, funzionalità. Per Phaidra è preferibile l'esportazione in in una cartella compressa (es. ZIP) contenente anche i file accessori. Esiste anche la possibilità di archiviazione in formato WARC (WebARChive), che tuttavia è

					complessa sia in creazione che in lettura
	XML		XML	Altro	Include formati di marcatura basati su XML comprensivi di DTD/Schema e fogli di stile XSD/XSL (TEI, DocBook...). La codifica dei caratteri può variare (UTF-8, UTF-16, ASCII...) e deve essere dichiarata esplicitamente
	EPUB		EPUB	Altro	Formato basato su XML. Non di largo uso, ma ben documentato. Non deve essere criptato o contenere limitazioni di accesso
	TXT		TXT	Altro	Il formato di testo più semplice e più supportato. La codifica dei caratteri può variare (UTF-8, UTF-16, ASCII...). È consigliato UTF-8
	ODT (ODF)		ODT (ODF)	Altro	Formato aperto. Specifiche interamente documentate. Preferire l'esportazione in PDF/A
	ODT (ODP)		ODP (ODF)	Altro	Formato aperto. Specifiche interamente documentate. Preferire l'esportazione in PDF/A
	DOCX (OOXML)		DOCX (OOXML)	Altro	Formato aperto. Specifiche interamente documentate. Preferire l'esportazione in PDF/A. Escludere: macro, file binari, riferimenti incrociati con file esterni
	PPTX (OOXML)		PPTX (OOXML)	Altro	Formato aperto. Specifiche interamente documentate. Preferire l'esportazione in PDF/A

Precisazioni

Per i PDF, Phaidra consente un doppio caricamento: PDF scaricabile e PDF “leggero” per una veloce visualizzazione nel browser (si veda la [Guida all'archiviazione](#)).

Il PDF generato dal *Phaidra Importer* non è di tipo PDF/A.

Immagine

Archiviazione Consigliato	Archiviazione Possibile	Web Consigliato	Web Possibile	Tipo di oggetto in Phaidra	Note
	JPEG	JPEG		Immagine	Di largo utilizzo, in particolare per la fruizione via web (file compressi, con perdita). Non adatto per la conservazione a lungo termine di immagini di qualità
	JPEG 2000 (JP2)		JPEG 2000 (JP2)	Altro	Standard coperto da brevetti; solo le specifiche della Parte1 sono completamente disponibili. Ne esistono varie implementazioni non necessariamente compatibili l'una con l'altra. Ancora non ampiamente utilizzato, ma con diffusione in crescita. Consente compressione senza perdita. Per l'archiviazione a lungo termine deve includere metadati descrittivi e tecnici
TIFF			TIFF	Immagine	Specifiche interamente documentate. È lo standard <i>de facto</i> per l'archiviazione delle immagini. Deve essere TIFF 6.0 non compresso, con ordine dei byte Intel (PC), e inclusione di metadati descrittivi e tecnici. Rispetto alle immagini con perdita, ad es. JPEG, si tratta di file pesanti
	PNG	PNG		Immagine	Formato compresso senza perdita, completamente aperto. Di ampio utilizzo in ambiti particolari

	PDF, PDF/A	PDF, PDF/A		Documento (PDF, TeX)	Formato di archiviazione ampiamente utilizzato. Ben supportato; specifiche interamente disponibili. PDF/A vieta l'uso di alcune funzioni del PDF che potrebbero essere difficili da rendere in futuro
SVG		SVG		Altro	Basato su XML. Formato aperto e interamente documentato, adatto all'archiviazione di immagini vettoriali
	DNG			Altro	È il formato RAW (negativo digitale) aperto; è interamente documentato
	RAW proprietari (NEF, CRW, 3FR...)			Altro	I formati RAW (negativi digitali) prodotti dalle fotocamere (CAM_RAW), non sono adatti all'archiviazione a lungo termine, in quanto proprietari (con l'eccezione di DNG). Tuttavia, purché venga archiviata anche una copia in formato TIFF, è utile conservare anche i RAW perché contengono i dati grezzi

Precisazioni

In generale per la distribuzione di immagini *raster* in Web, è da preferire il formato JPEG compresso, tuttavia per le immagini PNG che si vuole distribuire senza perdita di qualità utilizzare il formato nativo.

Il formato PDF è stato incluso qui – oltre che tra i formati testuali – perché può essere composto di sole immagini, ad esempio nel caso di libri digitalizzati. Si noti che non esiste unanimità di vedute sull'opportunità di un utilizzo del PDF/A per i PDF che contengono solamente immagini, per i

quali una conversione in PDF/A non ne migliora la conservazione a lungo termine (si legga ad esempio quanto scrive *The Open Preservation Foundation*)².

Per i PDF, Phaidra consente un doppio caricamento: PDF scaricabile e PDF leggero per una veloce visualizzazione nel browser (si veda la [Guida all'archiviazione](#)).

I PDF creati dal *Phaidra Importer* non sono di tipo PDF/A.

Nei PDF di immagini creati dal *Phaidra Importer* le immagini vengono in parte modificate (se non altro, si perdono i dati EXIF).

Per un confronto sui formati immagine si veda anche: FADGI, *Summary Table: Raster Still Images for Digitization: A Comparison of File Formats*, 2014, pag. 6 http://www.digitizationguidelines.gov/guidelines/FADGI_RasterFormatCompare_p3_20140417_r.pdf

2 <https://openpreservation.org/blog/2014/08/27/when-not-migrate-pdf-pdfa/>

Audio

Archiviazione Consigliato	Archiviazione Possibile	Web Consigliato	Web Possibile	Tipo di oggetto in Phaidra	Note
WAVE, Broadcast Wave (BWF)			WAVE, Broadcast Wave (BWF)	Audio	È uno standard <i>de facto</i> ³ , ben documentato. Per archiviazione, dovrebbe contenere solo audio con codifica Linear PCM bitstream (LPCM), non compresso. I file WAVE sono relativamente pesanti se paragonati ai formati con perdita. Broadcast Wave (BWF) sono WAVE file con metadati inclusi, e sono il formato preferito da IC, LOC, NAA, SIA, IASA, LAC e NARA (per NARA, assieme a FLAC) per l'archiviazione
	MP3	MP3		Audio	Formato originariamente coperto da brevetti (ora decaduti), specifiche pubbliche. Molto usato, anche grazie all'elevata comprimibilità. Compressione con perdita, quindi non adatto per archiviazione a lungo termine
FLAC			FLAC	Audio	Non utilizza il <i>player</i> di Phaidra

3 https://www.webarchive.org.uk/wayback/archive/20160101152346/http://www.jiscdigitalmedia.ac.uk/infokit/file_formats/audio-wrappers

	MPEG-4 AAC		MPEG-4 AAC	Altro	Molto usato, anche per l'alta comprimibilità. Compressione con perdita e coperto da brevetti, quindi non adatto per archiviazione a lungo termine. Non utilizza il <i>player</i> di Phaidra
	AIFF		AIFF	Altro	Per archiviazione dovrebbe contenere solo audio con codifica Linear PCM bitstream (LPCM), non compresso. Non utilizza il <i>player</i> di Phaidra

Precisazioni

In genere per l'archiviazione sono da preferire formati non compressi (sebbene la discussione sia aperta, in particolare per il formato FLAC) e alla risoluzione nativa piuttosto che con ricampionamento.

Si veda anche: CAVPP, *Target Audio and Video Specifications*, 2017 <https://calpreservation.org/wp-content/uploads/2017/03/CAVPP-File-Specs-2017.03.08.pdf>

4 <http://dericed.com/2013/flac-in-the-archives/>

Video

Un formato video è un oggetto complesso, che può essere visto come un contenitore di più file di diversa natura: audio, video, sottotitoli e altri. Tali file possono essere compressi con codifiche differenti, alcune più adatte alla conservazione a lungo termine altre alla fruizione web. Quindi, qui è necessario dettagliare tanto i formati del contenitore quanto le codifiche dei contenuti video e audio (in generale per la lunga conservazione l'audio sarà WAVE non compresso, o in alternativa FLAC); di conseguenza, le due tabelle che seguono vanno utilizzate congiuntamente⁵.

A differenza di quanto accade nella digitalizzazione audio, dove WAVE (normalmente nella variante BWA) è ampiamente considerato lo standard *de facto* per l'archiviazione, in ambito video non c'è consenso tra le istituzioni archivianti.

Per l'archiviazione video, storicamente si notano due tendenze contrapposte: le agenzie di diffusione televisiva, possedendo enormi quantità di riprese video nelle quali spesso il contenuto è più importante della qualità dell'immagine, tendono ad archiviare in formato compresso e con perdita (*lossy*), in modo da velocizzare le procedure e ridurre i costi di stoccaggio⁶. Viceversa le istituzioni (archivi, musei, biblioteche) che devono conservare il patrimonio culturale video del quale sono depositarie, preferiscono archiviare il materiale alla migliore qualità possibile, non compresso o, più frequentemente, con compressione *lossless*, cioè senza perdita di dati⁷.

Per l'archiviazione a lungo termine i grandi archivi internazionali suggeriscono svariati contenitori – AVI, MOV, MXF, Matroska – e svariate codifiche – JPEG2000, FFV1 o V210.

Secondo un'analisi dello studio comparativo di FADGI relativo ai maggiori progetti di digitalizzazione⁸, esisterebbero due differenti impostazioni, o “comunità”, quella dei grandi archivi e biblioteche nazionali più propense a utilizzare standard con la “S” maiuscola (MXF o MOV con codifica JPEG2000) e quelle degli specialisti, collocati principalmente in Europa, attivi nell'adozione di formati nati in progetti *open source* anche quando non ampiamente affermati o non ben documentati⁹ (Matroska con codifica FFV1)¹⁰.

5 Per una corrispondenza tra contenitori e formati di codifica supportati, si veda Wikipedia “*Comparison of video container formats*”

https://en.wikipedia.org/wiki/Comparison_of_video_container_formats

6 L'Ufficio Digital Learning e Multimedia dell'Università di Padova ad esempio, consiglia l'archiviazione nel formato originale, ma, nel caso il formato sia obsoleto, affiancato da una copia in un formato di largo utilizzo, anche compresso e con perdita (ad es. MP4 o Matroska, con codifica video H.264, e audio FLAC) (Marco Toffanin, comunicazione a voce)

7 http://download.das-werkstatt.com/pb/mthk/info/video/comparison_video_codecs_containers.html#lossy_vs_lossless

8 http://www.digitizationguidelines.gov/guidelines/video_reformatting_compare.html?loclr=blogsig

9 Il limite di non essere ben documentato presente in FFV1 è destinato a cadere a breve dal momento che le specifiche per FFV1 sono ora allo stato di “last call” grazie all'IETF cellar working group (<https://datatracker.ietf.org/doc/draft-ietf-cellar-ffv1/> Febbraio 2019)

10 <https://blogs.loc.gov/thesignal/2014/12/comparing-formats-for-video-digitization/>

Per la qualità video dei file destinati alla conservazione a lungo termine si prendano in considerazione anche la risoluzione¹¹, il *bitrate*¹² e tutti gli altri parametri in grado di conservare al meglio l'autenticità e l'integrità del video originale. In particolare, nel caso di conversione a nuovo formato, vanno mantenuti risoluzione e *bitrate* dell'originale.

In generale, per la conservazione è consigliata l'archiviazione nel formato originale, ma, nel caso in cui il formato sia obsoleto è consigliata anche l'archiviazione di una copia in un formato attualmente di largo utilizzo e idoneo allo scopo (quindi ben documentato, non compresso o compresso senza perdita, ecc.)

Per quanto riguarda la visualizzazione in Phaidra, il *player* attualmente utilizzato ([Video.js](#)) riproduce solo il formato MP4 con codifica video H.264 e codifica audio MP3 o AAC. È riprodotto anche il formato MOV a condizione di modificarne manualmente il *media type* da "video/quicktime" a "video/mp4" (nell'Editor dei metadati di Phaidra: scheda "Dati tecnici", campo "Formato"). In futuro è prevedibile che Phaidra adotterà soluzioni di *streaming* più efficienti, compatibili con un più ampio spettro di formati e di codifiche.

Contenitori

Archiviazione Consigliato	Archiviazione Possibile	Web Consigliato	Web Possibile	Tipo di oggetto in Phaidra	Note
AVI			AVI	Video	Formato molto diffuso, ben supportato e con specifiche interamente documentate. Per l'archiviazione, IC consiglia AVI non compresso (codifica non specificata). Anche LAC preferisce questo formato (AVI con <i>chroma subsampling</i> 4:2:2; non compresso), assieme a MXF e MOV

11 In ordine crescente, dalla risoluzione peggiore alla migliore: VHS, PAL, DVD, Blu-Ray (720p, 1080p), 4K, 8K (<https://datamanagement.univie.ac.at/en/about-phaidra/formats/background-knowledge/>).

12 In ordine crescente, dal *bitrate* peggiore al migliore: 1 Mbps (480p), 2-5 Mbps(720p), 4.5 Mbps (1080p), 9.8 Mbps (DVD), 40 Mbps (HB Blu-Ray) (<https://datamanagement.univie.ac.at/en/about-phaidra/formats/background-knowledge/>).

	MP4 (MPEG-4 part 14)	MP4 (MPEG-4 part 14)		Video	Contenitore molto diffuso e ben documentato. Solitamente usato con codifica video MPEG-H Part 2 (H.265/HEVC), MPEG-4 Part 10 (H.264/AVC) e MPEG-4 Part 2; MPEG-4 AAC è la sola codifica audio consentita. Per l'accesso, IC e CAVPP consigliano MPEG-4
Motion JPEG 2000 (MJP2 o MJ2)			Motion JPEG 2000 (MJP2 o MJ2)	Altro	Formato diffuso, con specifiche interamente documentate. Compressione <i>intraframe</i> . File pesanti. Per i suoi progetti NAA usa Motion JPEG 2000
MXF			MXF	Altro	Formato diffuso più a livello professionale che desktop, con specifiche interamente documentate. Nella riformattazione dei <i>videotape</i> per la conservazione il National Audio-Visual Conservation Center della LOC usa MXF (JPEG 2000 incluse in MXF). Anche LAC preferisce questo formato (sempre in combinazione con JPEG2000 compresso senza perdita), assieme ad AVI e MOV (4:2:2 <i>chroma subsampling</i> ; non compresso)
Matroska (MKV)			Matroska (MKV)	Video	Sviluppato come progetto <i>open source</i> . Formato aperto, di largo utilizzo e in rapida espansione. Supporta un gran numero di codifiche audio e video. Può contenere oggetti complessi. Formato consigliato da PREFORMA.

					Per la conservazione, Archivematica e SNA usano Matroska (con codifica video FFV1 e audio LPCM)
Quicktime (MOV)			Quicktime (MOV)	Video	Formato molto diffuso. Specifiche completamente disponibili. Può essere considerato come una variante di MP4 al quale ha dato origine e dal quale incorpora molti aggiornamenti ¹³ . Può contenere oggetti complessi. Usato da CAVPP per i master. Anche LAC preferisce questo formato (MOV con codifica 4:2:2 non compressa), assieme a MXF e AVI
	OGG		OGG	Altro	Formato aperto, sviluppato dal progetto <i>open source</i> Xiph. Diffusione limitata. Può incorporare differenti codifiche audio e video
DPX			DPX	Altro	Formato aperto. È considerato uno standard per la conversione in digitale di alta qualità a partire da pellicole cinematografiche. Di norma non include audio, che viene salvato separatamente (di solito in WAVE). Utilizzato da Motion Picture, Broadcasting, and Recorded Sound Division (LOC), NARA e LAC

13 Sull'interdipendenza tra i formati MP4 e MOV si ringrazia Jérôme Martinez (comunicazione personale)

Formati di codifica video¹⁴

Archiviazione Consigliato	Archiviazione Possibile	Web Consigliato	Web Possibile	Note
	H.264 (anche AVC) (MPEG-4 part 10)	H.264 (anche AVC) (MPEG-4 part 10)		Formato compresso, con o senza perdita ¹⁵ , di largo utilizzo soprattutto per la fruizione web. Protetto da vari brevetti. Per accesso web LOC e CAVPP usano MPEG-4_AVC
	H.265 (anche HEVC) (MPEG-H Part 2)	H.265 (anche HEVC) (MPEG-H Part 2)		Formato successore di H.264, a maggiore efficienza, ma estremamente esigente in termini di risorse CPU. Coperto da numerosi brevetti. Diffusione in crescita.
JPEG 2000 (JP2)			JPEG 2000 (JP2)	Standard coperto da brevetti, compressione (<i>intraframe</i>) con o senza perdita. Esigente in termini di risorse CPU. Supportato da pochi software. Adottato da importanti archivi audio-video. Ne esistono varie implementazioni non necessariamente compatibili l'una con l'altra. Usato da NAA per i suoi progetti (in contenitore Motion JPEG 2000), dal National Audio-Visual Conservation Center della LOC nella riformattazione dei <i>videotape</i> per la conservazione (in contenitore MXF) e da LAC
	Theora		Theora	Formato aperto, sviluppato dal progetto <i>open source</i> Xiph. Compressione con perdita. Diffusione limitata
FFV1			FFV1	Per la conservazione, Archivematica usa FFV1 in contenitore Matroska, OM usa FFV1 in contenitore AVI. Su FFV1 in Matroska, si veda anche FIAF ¹⁷

¹⁴ Erroneamente a volte ci si riferisce ai formati di codifica col termine inglese *codec* (che sta per “coder-decoder”). Per una distinzione terminologica tra “formato di codifica” e “*codec*” si veda Wikipedia: [https://en.wikipedia.org/wiki/Video_coding_format#Distinction_between "format" and "codec"](https://en.wikipedia.org/wiki/Video_coding_format#Distinction_between_%22format%22_and_%22codec%22)

¹⁵ In realtà il formato di codifica non è ottimizzato per una compressione senza perdita e la maggior parte dei *player* non sono in grado di leggere H.264 compressi senza perdita (comunicazione personale di Peter Bubestinger-Steindl).

	Dirac		Dirac	Formato di codifica compresso, con o senza perdita, non coperto da brevetti. Diffusione limitata. Esigente in termini di risorse CPU. Sviluppato dalla BBC.
<i>V210, YUY2, UYVY e altri (tutti con 4:2:2 chroma subsampling)</i>			<i>V210, YUY2, UYVY e altri (tutti con 4:2:2 chroma subsampling)</i>	Formati ben documentati e ben supportati, non compressi, utilizzati da numerosi archivi. Risultano in file di grandi dimensioni. V210 è usato da CAVPP per i master. YUY2 è usato da NARA per la riformattazione di <i>videotape</i>

Precisazioni

La Figura 1 riassume graficamente la diffusione di contenitori e codifiche video consigliati per la lunga conservazione da parte di alcuni importanti archivi e progetti nel mondo.

Si può notare come alcuni archivi lascino un ampio spettro di scelta, anche includendo formati considerati “a rischio” (ad es. WMV), mentre altri restringano la scelta praticamente a un singolo formato (ad es. PREFORMA con formato di codifica FFV1 incluso in contenitore MKV, oppure LOC con formato di codifica JPEG2000 incluso in contenitore MXF). Nella figura dei formati consigliati è stato inserito anche NARA, nonostante questi archivi, per i formati video, prevedano solo formati “accettati” e nessuno “preferito”.

Tra i contenitori, accanto ai “classici” MOV e AVI, anche MXF e Motion JPEG2000 sono ampiamente utilizzati, e pure l'*open source* MKV comincia ad avere una significativa base di utilizzo per la conservazione in ambito professionale.

Tra i formati di codifica, oltre a quelli non compressi (4:2:2 e altri) spiccano l'utilizzo di JPEG 2000 e del formato *open source* FFV1.

Si noti anche come alcuni contenitori “tradizionali” – in particolare AVI e MOV – siano utilizzati in associazione a una gran varietà di codifiche, mentre i contenitori di più recente affermazione vengano usati in configurazioni più definite (MXF con JPEG2000 e MKV con FFV1).

16 https://www.webarchive.org.uk/wayback/archive/20160101152356/http://www.jiscdigitalmedia.ac.uk/infokit/file_formats/video-codecs

17 https://www.fiafnet.org/images/tinyUpload/E-Resources/Commission-And-PIP-Resources/TC_resources/FFV1_and_Matroska_reading_list.pdf

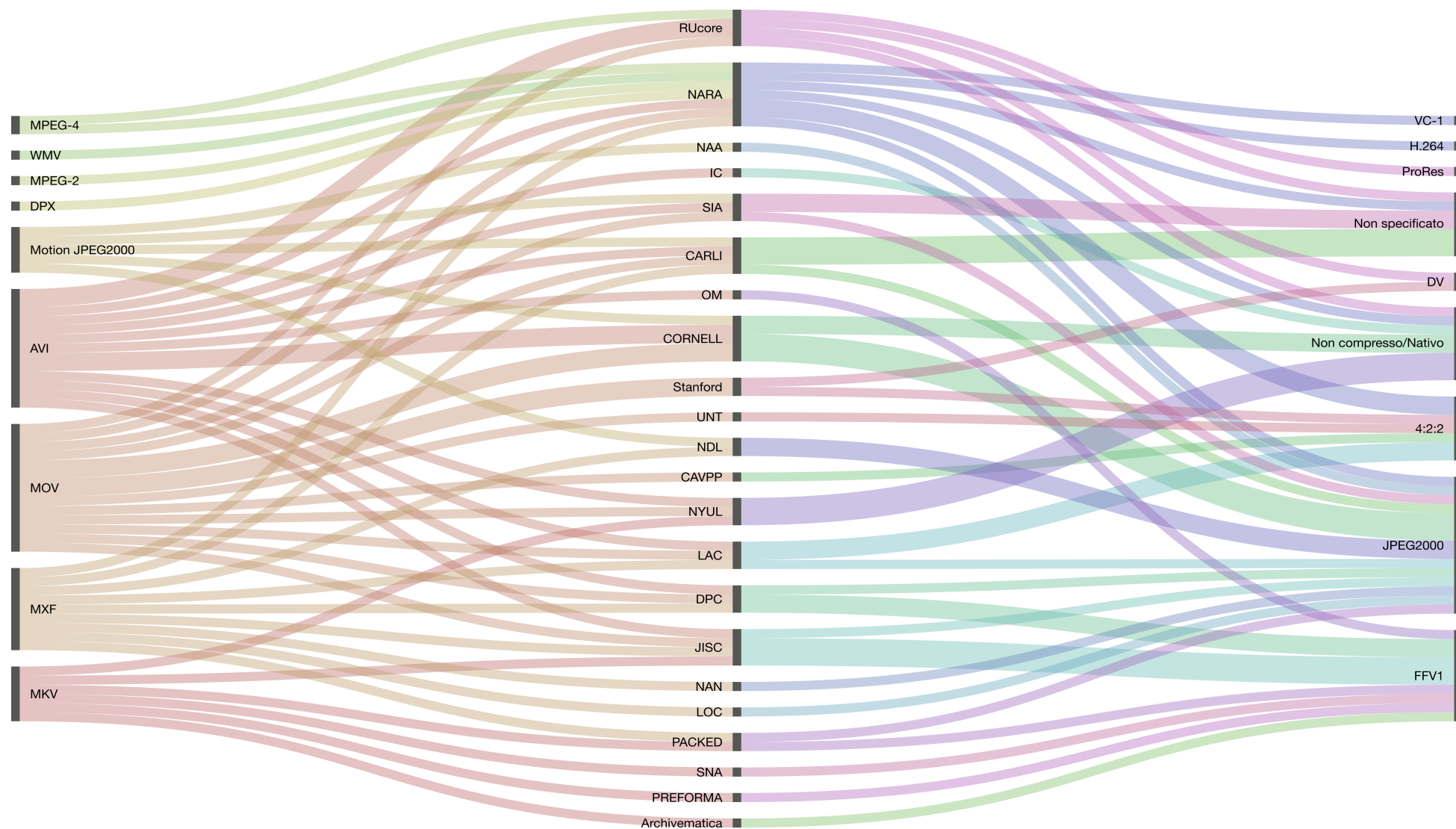


Figura 1: Diffusione di contenitori (incolonnati a sinistra) e codifiche (a destra) video consigliati per la lunga conservazione da parte di alcuni importanti archivi e progetti nel mondo (al centro)

Un utile strumento per determinare i dati tecnici di un file video (contenitore, codifica, bitrate, frame rate, spazio colore, profondità di bit e altro) è MediaInfo, disponibile per Windows, MacOS e Linux: <https://mediaarea.net/en/MediaInfo>. Gli stessi sviluppatori di MediaInfo stanno lavorando a uno strumento di analisi più approfondita, MediaTrace (<https://mediaarea.net/MediaTrace>) disponibile ad esempio in MediaConch (<https://mediaarea.net/MediaConch>).

Data la complessità dell'argomento, per approfondimenti sui formati dei file video destinati alla conservazione, si rimanda all'estesa documentazione presente in rete:

- FADGI, *Digital File Formats for Videotape Reformatting – Part 5. Narrative and Summary Tables*, 2014. A pag. 16-17 due tabelle, una per i wrapper e una per le codifiche http://www.digitizationguidelines.gov/guidelines/FADGI_VideoReFormatCompare_p5_20140908.pdf
- FADGI, *Digital File Formats for Videotape Reformatting*, 2014
http://www.digitizationguidelines.gov/guidelines/video_reformatting_compare.html
- FADGI, *Guidelines: MXF Application Specification*, 2017 http://www.digitizationguidelines.gov/guidelines/MXF_app_spec.html
- FADGI, *Creating and Archiving Born Digital Video*, 2014 http://www.digitizationguidelines.gov/guidelines/FADGI_BDV_p1_20141202.pdf http://www.digitizationguidelines.gov/guidelines/FADGI_BDV_p2_20141202.pdf
http://www.digitizationguidelines.gov/guidelines/FADGI_BDV_p3_20141202.pdf
http://www.digitizationguidelines.gov/guidelines/FADGI_BDV_p4_20141202.pdf
- NDL, *File formats*, pag 13 e segg. <http://digitalpreservation.fi/files/File-Formats-1.6.1-en.pdf>
- Wikipedia, *Comparison of video container formats* https://en.wikipedia.org/wiki/Comparison_of_video_container_formats
- Harvard Library Digital Preservation Program, *Video Format Matrix*, 2016
<https://wiki.harvard.edu/confluence/display/digitalpreservation/Video+Formats>
- IASA, *Guidelines for the Preservation of Video Recordings IASA-TC 06*, 2018 https://www.iasa-web.org/sites/default/files/publications/IASA-TC_06-B_20180518.pdf https://www.iasa-web.org/sites/default/files/publications/IASA-TC_06-B-app_20180520.pdf
- CAVPP, *Target Audio and Video Specifications*, 2017 <https://calpreservation.org/wp-content/uploads/2017/03/CAVPP-File-Specs-2017.03.08.pdf>

- AMIA, *Digital Formats - Video Part One*, 2017 <https://www.youtube.com/watch?v=BkYLEsjldK4>
- AMIA, *Digital Formats - Video Part Two*, 2017 <https://www.youtube.com/watch?v=-TOJp-AL8Z4>
- Stanford Media Preservation Lab, *Capture specs*, <https://library.stanford.edu/research/digitization-services/labs/stanford-media-preservation-lab/capture-specs>
- RUcore, *Video and Moving Image Objects: Recommended Minimum Standards For Archival and Presentation Datastreams*, 2015 <http://odin.rutgers.edu/standards/2015/Video%20Object%20Standards%20Analysis-2015.pdf>
- ALCTS, *Minimum Digitization Capture Recommendations*, 2013 <http://www.ala.org/alcts/resources/preserv/minimum-digitization-capture-recommendations>
- George Blood [for the LOC], *Refining Conversion Contract Specifications: Determining Suitable Digital Video Formats for Medium-term Storage*, 2011 http://www.digitizationguidelines.gov/audio-visual/documents/IntrmMastVidFormatRecs_20111001.pdf
- New York University Libraries, *Digitizing Video for Long-Term Preservation: An RFP Guide and Template*, 2013 <http://memoriav.ch/wp-content/uploads/2014/07/VARRFP.pdf>
- CARLI (Consortium of Academic and Research Libraries in Illinois), *Guidelines for the Creation of Digital Collections, Digitization Best Practices for Moving Images*, 2017 https://www.carli.illinois.edu/sites/files/digital_collections/documentation/guidelines_for_video.pdf
- FIAF, FFV1 and Matroska reading list https://www.fiafnet.org/images/tinyUpload/E-Resources/Commission-And-PIP-Resources/TC_resources/FFV1_and_Matroska_reading_list.pdf
- LAC, *Guidelines on File Formats for Transferring Information Resources of Enduring Value*, 2014 <https://www.bac-lac.gc.ca/eng/services/government-information-resources/guidelines/Pages/guidelines-file-formats-transferring-information-resources-enduring-value.aspx>
- LAC presenta una lista di coppie “contenitore/codifica” preferite o accettabili a pag. 14 di <https://www.bac-lac.gc.ca/eng/services/government-information-resources/guidelines/Documents/file-formats-irev.pdf>
- P. Bubestinger-Steindl, H. Lewetz, M. Jaks, *Comparing Video Codecs and Containers for Archives*, 2015 http://www.av-rd.com/knowhow/video/comparison_video_codecs_containers.html

- P. Bubestinger-Steindl, *Risk Assessment Considerations: Using FFV1 for Preservation*, 2016 http://www.av-rd.com/knowhow/video/risk_assessment.html
- eCommons: Cornell's Digital Repository, *Recommended File Formats*, 2018 <https://guides.library.cornell.edu/ecommons/formats>
- Smithsonian Institution Archives, *Recommended Preservation Formats for Electronic Records* <https://siarchives.si.edu/what-we-do/digital-curation/recommended-preservation-formats-electronic-records>
- Reto Kromer, *Matroska and FFV1: One File Format for Film and Video Archiving?*, 2017 https://www.fiafnet.org/images/tinyUpload/Publications/Journal-Of-Film-Preservation/Matroska-and-FFV1_Kromer_JFP96.pdf
- National Archives of the Netherlands, *Preferred formats National Archives of the Netherlands: in view of sustainable accessibility*, 2016 <https://www.nationaalarchief.nl/sites/default/files/field-file/National%20Archives%20of%20the%20Netherlands%20preferred%20and%20acceptable%20formats.pdf>
- Reto Kromer, *Matroska and FFV1: One File Format for Film and Video Archiving?*, in *Journal of Film Preservation*, n. 96 (April 2017) https://retokromer.ch/publications/JFP_96.html

Dati strutturati

Archiviazione Consigliato	Archiviazione Possibile	Web Consigliato	Web Possibile	Tipo di oggetto in Phaidra	Note
CSV		CSV		Altro	Formato aperto. Standard <i>de facto</i> . Uno dei formati preferiti da LOC
TSV		TSV		Altro	Formato aperto. Uno dei formati preferiti da LOC
	ODS (ODF)		ODS (ODF)	Altro	Formato aperto. Specifiche interamente documentate. Preferire l'esportazione in PDF/A, o CSV. Vanno comunque esclusi: macro, file binari, riferimenti incrociati con file esterni
	XLXS (OOXML)		XLXS (OOXML)	Altro	Formato aperto. Preferire l'esportazione in PDF/A o CSV. Vanno comunque esclusi: macro, file binari, riferimenti incrociati con file esterni
JSON		JSON		Altro	Formato aperto. Tipicamente un formato di interscambio per i dati. Uno dei formati preferiti da LOC
XML		XML		Altro	Formato aperto. Si può dichiarare la codifica nell'Editor dei metadati (Dati tecnici → Requisiti per l'utilizzo dell'oggetto). Es. UTF-8 e UTF-16 (con BOM), US-ASCII, ISO 8859...
	PDF/A	PDF/A		Documento (PDF, TeX)	Formato aperto. Formato di archiviazione ampiamente utilizzato, che vieta l'uso di alcune funzioni del PDF che potrebbero essere difficili da

					rendere in futuro. Ben supportato; specifiche interamente disponibili
--	--	--	--	--	---

Precisazioni

I file di dati e i database devono essere trasferiti come file piatti o come tabelle rettangolari, ovvero come *array* bidimensionali, elenchi o tabelle. I dati strutturati devono essere trasferiti insieme ai file associati necessari per verificare la validità dei dati, ad es. DTD, schemi e dizionari di dati.

Ringraziamenti

Si ringraziano Marco Toffanin¹⁸ e Antonio Zanonato¹⁹ per la consulenza sui formati video.

Si ringraziano Jérôme Martinez²⁰ e Peter Bubestinger-Steindl²¹ per la revisione delle sezioni su audio e video.

18 Ufficio digital learning e multimedia dell'Università di Padova

19 Dipartimento dei Beni Culturali dell'Università di Padova

20 https://archive.fosdem.org/2018/schedule/speaker/jerome_martinez/

21 <http://www.preforma-project.eu/advisory-board.html>

Abbreviazioni usate

AGID = Agenzia per l'Italia digitale <https://www.agid.gov.it/>

ALCTS = Association for Library Collections & Technical Services <http://www.ala.org/alcts/resources/preserv/minimum-digitization-capture-recommendations>

AMIA = Association of Moving Image Archivists <https://amianet.org/>

Archivematica https://wiki.archivematica.org/Media_type_preservation_plans

CAVPP = California Audiovisual Preservation Project <https://calpreservation.org/>

CORNELL = eCommons: Cornell's Digital Repository <https://guides.library.cornell.edu/eccommons>

CPP = California Preservation Program <https://calpreservation.org>

DCC = Digital Curation Centre <http://www.dcc.ac.uk/>

DPC = Digital Preservation Coalition <https://www.dpconline.org/>

FADGI = Federal Agencies Digitization Guidelines Initiative <http://www.digitizationguidelines.gov/>

FIAF = International Federation of Film Archives <https://www.fiafnet.org>

FIAT = Fédération Internationale des Archives de Télévision / The International Federation of Television Archives (FIAT/IFTA)
<http://fiatifta.org/index.php/about/>

IASA = International Association of Sound and Audiovisual Archives (IASA) <https://www.iasa-web.org/>

IC = Internet Culturale <http://www.internetculturale.it/it/1131/linee-guida-e-standard>

JISC = Joint Information Systems Committee Digital Media https://www.webarchive.org.uk/wayback/archive/20160101151358/http://www.jiscdigitalmedia.ac.uk/infokit/file_formats/digital-file-formats

LAC = Library and Archives Canada <http://www.bac-lac.gc.ca/eng/Pages/home.aspx>

LOC = Library of Congress <https://www.loc.gov/preservation/digital/formats/fdd/descriptions.shtml>

NAA = National Archives of Australia <http://www.naa.gov.au/information-management/managing-information-and-records/preserving/long-term-file-formats.aspx>

NAN = National Archives of the Netherlands <https://www.nationaalarchief.nl>

NARA = The U.S. National Archives and Records Administration <https://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html>

NDL = National Digital Library's Digital preservation service (Finland) <http://digitalpreservation.fi/en/specifications>

OM = Österreichische Mediathek <https://www.mediathek.at/>

PREFORMA – PREservation FORMAts for culture information and e-archives <http://www.preforma-project.eu/media-type-and-standards.html>

PRONOM – Digital Preservation Department of the UK National Archives <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>

SIA = Smithsonian Institution Archives <https://siarchives.si.edu/what-we-do/digital-curation/recommended-preservation-formats-electronic-records>

SMPTE = Society of Motion Picture and Television Engineers <https://www.smpte.org/>

SNA = Swedish National Archives <https://riksarkivet.se/startpage>

UNT = University of North Texas Libraries, Digital projects unit <https://library.unt.edu/digital-projects-unit/standards/>

UW = Universität Wien <https://datamanagement.univie.ac.at/en/about-phaidra/formats/>

Wikipedia – List of filename extensions https://en.wikipedia.org/wiki/List_of_filename_extensions – List of file formats https://en.wikipedia.org/wiki/List_of_file_formats#Video