

Non solo TIFF e JPEG: riflessioni sui formati elettronici per la digitalizzazione del patrimonio archivistico e librario e sul nuovo standard PDF/R

«DigItalia» 2-2022
DOI: 10.36181/digitalia-00050

Stefano Allegrezza

Università degli Studi di Bologna

Il contributo intende fornire una riflessione sulla scelta dei formati elettronici da utilizzare per l'acquisizione delle immagini in un progetto di digitalizzazione del patrimonio archivistico e librario. Fino ad oggi la scelta è caduta quasi sempre sul formato TIFF per quanto riguarda la produzione di immagini master mediante scansione (o, eventualmente, su uno dei formati RAW nel caso di acquisizione mediante macchina fotografica) e sul formato JPEG per quanto riguarda la produzione di immagini derivate.

Tuttavia molti si domandano se queste scelte siano ancora corrette e se non vi siano altre possibilità che, da una parte, forniscano migliori garanzie per quanto riguarda la conservazione a lungo termine e, dall'altra, consentano una maggiore fruibilità. A questo proposito merita una attenta riflessione l'adozione del nuovo formato PDF/R, che ha ottenuto recentemente il riconoscimento come standard ISO, e che si propone come sostituto sia del formato TIFF che del formato JPEG. Trattandosi di un profilo del formato PDF ha ottime caratteristiche di fruibilità; inoltre, ha anche eccellenti caratteristiche di conservabilità a lungo termine, grazie alla sua conformità alle specifiche del PDF/A. Nel contributo verrà dapprima proposta una panoramica delle indicazioni fornite dalle principali linee guida nazionali in materia, poi verranno analizzate le alternative al TIFF e al JPEG come formati per la produzione di file master e file derivati, rispettivamente; infine, verrà analizzato il nuovo formato PDF/R mettendone in evidenza le caratteristiche che lo propongono come formato di elezione per i progetti di digitalizzazione del patrimonio archivistico e librario.

Introduzione

Una delle decisioni da prendere quando si avvia un progetto di digitalizzazione del patrimonio archivistico e librario è quella relativa ai formati elettronici nei quali devono essere prodotte sia la copia conservativa (file master) che le copie derivate. Nel caso di materiale cartaceo, la scelta relativa al file master ricade quasi sempre sul formato TIFF (a singola pagina o multi pagina) nel caso in cui

l'acquisizione avvenga mediante scanner¹, oppure sul formato RAW (o eventualmente sul DNG) se l'acquisizione avviene mediante fotocamera digitale. Per i formati derivati viene scelto tipicamente il formato JPEG con differenti livelli di compressione; talvolta si impiegano anche altri formati, come il JPEG2000, il PDF (a pagina singola o multi pagina), il BMP o il PNG. Le varie linee guida propongono il TIFF come formato di elezione per la produzione delle immagini master nel caso di acquisizione mediante scansione, e il JPEG come formato di elezione per la produzione di immagini derivate, ma molti hanno cominciato a chiedersi se si tratti effettivamente di scelte corrette, soprattutto dopo il recente riconoscimento come standard ISO 23504:2020 del PDF/R, un profilo del formato PDF specificatamente pensato per l'acquisizione delle immagini nei progetti di digitalizzazione. Con questo articolo si cercherà di fare un po' di chiarezza e proporre qualche riflessione sull'argomento².

Le indicazioni fornite dalle principali linee guida nazionali

Il tema della scelta dei formati elettronici da utilizzare per la produzione delle immagini acquisite durante un processo di digitalizzazione è presente in tutte le linee guida, sia a livello italiano che internazionale³. Per semplicità ci si limiterà in questa sede a prendere in esame le indicazioni fornite dalle principali linee guida e raccomandazioni nazionali.

¹ Per tipologie e caratteristiche degli scanner si rinvia alle *Linee guida per la digitalizzazione del patrimonio culturale* del Ministero della Cultura, versione giugno 2022, disponibili all'indirizzo <<https://digitalibrary.cultura.gov.it/linee-guida>>. In particolare si vedano il paragrafo D.1.A.1 e seguenti.

² Cfr. l'articolo *Which File Format Should I Use for Scanned Documents?*, pubblicato all'indirizzo <<https://www.datamation.com/resource/file-format-scanned-documents-pdf-tiff>>, 5 novembre 2019.

³ In questo saggio si limiterà l'analisi al solo panorama italiano. Per un approfondimento sulle risorse internazionali, si faccia riferimento alle seguenti linee guida: UK National Archives, *Digitisation at The National Archives*, 2016, <<https://cdn.nationalarchives.gov.uk/documents/information-management/digitisation-at-the-national-archives.pdf>>; FADGI (Federal Agencies Digital Guidelines Initiative), <<http://www.digitizationguidelines.gov>>, in particolare: *Technical Guidelines for Digitizing Cultural Heritage Materials Creation of Raster Image Files*, 2016; METAMORFOZE, Netherlands' national programme for the preservation of paper heritage, *Metamorfoze Preservation Imaging Guidelines*, <https://www.metamorfoze.nl/sites/default/files/documents/Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf>; National Archives of Australia, *Guidelines for handling, preparing and digitising archival paper and printed materials*, 2022, <<https://www.naa.gov.au/guidelines-handling-preparing-and-digitising-archival-paper-and-printed-materials>>; ISO/TR 13028:2010, Information and documentation - Implementation guidelines for digitization of records; ATHENA WP3, *Digitisation: standards landscape for European museums, archives, libraries*, 2009; Progetto Minerva (Ministerial Network for Valorising Activities in Digitisation), *Manuale di buone pratiche per la digitalizzazione del patrimonio culturale. Versione 1.3*, a cura del Gruppo di lavoro n. 6 del progetto Minerva, Identificazione delle buone pratiche e dei centri di competenza, 2004, <https://www.minervaeurope.org/structure/workinggroups/goodpract/document/buonepratiche1_3.pdf>.

Innanzitutto, la pagina “Linee guida e standard per la realizzazione di progetti di digitalizzazione” pubblicata sul portale Internet Culturale⁴ prevede per la scelta dei formati per immagini le seguenti indicazioni: a) per la copia master, destinata alla conservazione come copia di sicurezza o a fini commerciali dai detentori dei diritti, «il formato TIFF 6.0 non compresso ad alta risoluzione, 600 dpi ottici, profondità di colore di 24 bit RGB per formato inferiore o uguale ad A4; e con una risoluzione di 400 dpi ottici, profondità di colore di 24 bit RGB per formato decisamente superiore ad A4»; b) per la copia web, un formato compresso a bassa risoluzione, identificato nel formato «JPEG 150 dpi ottici, profondità di colore di 24 bit RGB»⁵; c) per la copia Intranet, qualora se ne ravvisi la necessità, un formato compresso a media risoluzione, identificato nel formato «JPEG 300 dpi ottici, profondità di colore di 24 bit RGB».

Indicazioni simili vengono fornite dalle “Linee guida per i progetti di digitalizzazione del patrimonio archivistico e librario” a cura dell’Ufficio Nazionale per i beni culturali ecclesiastici e l’edilizia di culto della Conferenza Episcopale Italiana (versione 8 novembre 2019, aggiornate a novembre del 2021)⁶. Dopo aver ricordato che «la scelta del formato è generalmente diversa per l’immagine Master (o di conservazione) e per le immagini derivate», esse specificano che generalmente «si utilizza per l’immagine Master il formato TIFF non compresso» oppure con una compressione senza perdita di qualità, LZW per esempio; «per le immagini derivate (ad esempio, quelle per la consultazione) i formati con compressione, e tra questi il più utilizzato è il formato JPEG». Inoltre, nel caso di fascicoli o volumi rilegati le linee guida suggeriscono che «può essere opportuno creare, a partire dai file prodotti per le singole pagine/fogli, un unico file per la consultazione. In questo caso il formato consigliato è il formato PDF (Portable Document Format), formato che consente di rappresentare documenti in modo indipendente dal software e dall’hardware utilizzati. Il formato PDF è tra i più diffusi e supportati». Riguardo l’uso del formato PDF le linee guida precisano che «l’utilizzo di tale formato va inteso come aggiuntivo rispetto alla produzione dell’immagine master e derivata».

Anche le “Linee guida per i progetti di digitalizzazione” redatte dalla Regione

NARA - U.S. National Archives and Records Administration, *Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images*, 2004, <<https://www.archives.gov/files/preservation/technical/guidelines.pdf>>; IFLA, *Guidelines for digitization projects for collections and holdings in the public domain, particularly those held by libraries and archives*, 2008, <<https://www.ifla.org/files/assets/preservation-and-conservation/publications/digitization-projects-guidelines.pdf>>.

⁴ Si veda la pagina “Linee guida e standard” all’indirizzo

<<https://www.internetculturale.it/it/1131/linee-guida-e-standard>>.

⁵ Le linee guida specificano che «il fattore di ridimensionamento può essere ridefinito in fase prototipale, in funzione di un’agevole consultabilità, a seconda della qualità del materiale (inchiostri, qualità della scrittura, della stampa, corpo dei caratteri), anche in relazione alle dimensioni dell’originale».

⁶ Le linee guida sono disponibili all’indirizzo: <<https://bce.chiesacattolica.it/2021/11/26/linee-guida-per-i-progetti-di-digitalizzazione-del-patrimonio-archivistico-e-librario>>.

Piemonte e dal CSI Piemonte⁷, propongono considerazioni analoghe: dopo aver distinto «le diverse classi di file digitali a seconda dell'utilizzo a cui sono destinati» tra "master" (destinati all'archiviazione e conservazione) e "dissemination" (destinati alla divulgazione e alla consultazione sul web), specificano che per la prima categoria (master) è necessario scegliere formati che permettano una qualità molto alta dell'immagine, siano senza perdita di dati nel caso in cui vengano compressi e garantiscano una certa stabilità in modo da facilitare le pratiche di conservazione; per quanto riguarda la seconda categoria, la scelta deve essere orientata verso formati che permettano una media qualità dell'immagine, siano ampiamente supportati da browser, non siano proprietari, non richiedano software specifici per la visualizzazione. Da tali principi generali discende che, per i file master occorre scegliere immagini ad alta risoluzione in formato TIFF non compresso (o con compressione LZW), con profondità delle riprese a colori (24 bit) o in toni di grigio (8 bit)⁸. Per quanto riguarda, invece, «i file "dissemination", occorre scegliere il formato JPEG compresso con valori indicativi di risoluzione finale pari a 300 dpi per dimensioni fino a A4, 150 dpi per dimensioni fino a A3 e 75 dpi per dimensioni superiori a A3. Anche queste linee guida specificano che è possibile utilizzare il formato PDF per «la consegna di file in formato .PDF contenenti le immagini a bassa risoluzione degli oggetti considerati come "insieme" (volume, insieme archivistico, quindi un file per ogni opera)», che è richiesta in caso di progetti di digitalizzazione di monografie e/o documenti d'archivio, per agevolarne la consultazione.

Le "Linee Guida sulla digitalizzazione" (revisione ottobre 2017) elaborate dal Sistema bibliotecario di Ateneo dell'Università di Padova (comunemente note come Linee guida Phaidra), allargano la rosa di formati elettronici per la produzione di file master. Infatti, esse specificano che, «a seconda delle caratteristiche dello strumento di acquisizione utilizzato i file master possono essere di 2 tipi: immagini TIFF e immagini RAW (i cosiddetti negativi digitali), in uno dei vari formati proprietari della fotocamera, ad esempio NEF per le Nikon o CR2 per le Canon. Nel caso il formato dei master fosse un formato RAW conviene farne una copia in formato TIFF 6.0 non compresso, per garantirsi nel tempo la leggibilità nei software di uso comune. Questi TIFF devono essere fedeli all'originale RAW e quindi non essere rielaborati, tranne che per la correzione dei colori, operazione che si compie con maggiore efficacia e sicurezza nei file RAW». Per quanto riguarda i file derivati, le linee guida Phaidra suggeriscono l'utilizzo o del formato TIFF 6.0, non compresso "riscalato" a 2.400 pixel sul lato lungo, oppure del formato JPEG ottenuto utilizzando una compressione più o meno elevata a seconda che si intenda ottenere delle immagini

⁷ Le linee guida sono disponibili all'indirizzo: <https://www.memora.piemonte.it/cms/sites/default/files/2021-09/EcosistemaBeniCulturali_LineeGuida_digitalizzazione_V01.pdf>.

⁸ Le "Linee guida per i progetti di digitalizzazione" stabiliscono che «i valori indicativi di risoluzione finale effettiva (dimensione dell'immagine uguale al documento originale) sono 600 dpi per dimensioni fino a A4, 300 dpi per dimensioni fino a A3 e 200 dpi per dimensioni superiori a A3».

JPEG di alta, media o bassa qualità. L'ultima versione delle suddette linee guida (revisione luglio 2022) si allinea alle indicazioni fornite dalle "Linee guida per la digitalizzazione del patrimonio culturale" del Piano nazionale di digitalizzazione del patrimonio culturale (PND) (vedi *infra*), introducendo la distinzione tra il flusso di lavoro per acquisizione digitale da fotocamera e quello per acquisizione digitale da scanner. Nel primo caso si suggerisce l'impiego del formato RAW – definito anche "negativo digitale" – che è «normalmente il miglior formato prodotto dalle fotocamere»⁹. Secondo queste linee guida «i master in formato raw prodotti dalle fotocamere vanno conservati così come escono dallo strumento di acquisizione, cioè senza ulteriori elaborazioni. Tuttavia, essendo in un formato proprietario, non adeguato per la conservazione a lungo termine, vanno convertiti nel formato raw aperto DNG, non compresso. I file nel formato raw proprietario a questo punto vanno cancellati». Le linee guida specificano, altresì, che «le impostazioni in macchina fotografica al momento della ripresa dovranno garantire una adeguata qualità di colore (ad esempio spazio colore ProPhoto RGB o Adobe RGB e profondità di colore a 16 o 8 bit per canale)»¹⁰. Per quanto riguarda l'acquisizione digitale mediante scanner, le linee guida precisano che «di solito il miglior formato prodotto dagli scanner è il formato TIFF». A seconda della tipologia di materiale da acquisire occorre impostare i parametri corretti in termini di risoluzione, profondità di colore, spazio colore¹¹; in ogni caso il formato consigliato è il TIFF 6.0 non compresso.

⁹ A questo proposito le linee guida spiegano che «ogni casa produttrice ha il suo proprio tipo di formato raw proprietario, con proprie specifiche, che dà luogo a file con estensione caratteristica (ad esempio, .NEF per le Nikon, .CR2 o .CR3 per le Canon). Esiste un formato raw aperto, il formato DNG, ma normalmente non è supportato dalle fotocamere». Le linee guida sono disponibili all'indirizzo: <<https://phaidra.cab.unipd.it/static/linee-guida-digitalizzazione.pdf>>.

¹⁰ Per "spazio colore" si intende la gamma dei colori che possono essere riprodotti da un dispositivo (ad esempio, a schermo o a stampa). La gamma più ampia di colori è rappresentata dallo spettro visibile in natura, contenente tutti i colori che possono essere percepiti dall'occhio umano, ma i dispositivi di acquisizione (fotocamere, scanner) possono avere più o meno limitazioni nel riprodurre l'intera gamma cromatica, e così pure i dispositivi di riproduzione (monitor, stampanti), quindi utilizzano spazi colore limitati. Tra gli spazi colore più utilizzati per la finalità di acquisizione digitale, si ricorda l'sRGB, nato nel 1996 dalla collaborazione di Microsoft e HP, che è lo spazio colore più ristretto – rappresentando solo il 35% dei colori visibili dall'occhio umano – ma comunque sufficiente per la diffusione delle immagini in Internet e per la visualizzazione sui monitor (che in genere, a differenza delle stampanti, non sono in grado di rappresentare uno spazio colore più ampio); l'Adobe RGB (1998), che è in grado di rappresentare il 50% dei colori; il ProPhoto RGB, che è in grado di rappresentare il 90% dei colori. In un progetto di digitalizzazione occorre specificare qual è lo spazio di colore utilizzato: questa informazione viene resa attraverso il "profilo colore", tipicamente costituito da un piccolo file con un'estensione .ICC o .ICM. Alcuni formati per immagini consentono di integrare i profili del colore all'interno delle immagini stesse per specificare la gamma cromatica dei dati così da assicurare che gli utenti vedano gli stessi colori su dispositivi diversi.

¹¹ Le linee guida distinguono tra materiale grafico (fotografie, stampe, disegni, dipinti, manifesti, mappe, carte geografiche ecc.); libri, riviste e manoscritti, rari o di pregio (ad esempio illustrati o dipinti) oppure con scarsa leggibilità (caratteri sbiaditi, basso contrasto, note a margine a matita, macchiati); libri, riviste, manoscritti, dattiloscritti e ciclostilati, non rari, né di pregio, ben leggibili; negativi, diapositive in bianco e nero; negativi, diapositive a colori.

Anche la Soprintendenza Archivistica e Bibliografica della Campania ha fornito indicazioni in merito con le Linee guida “Progetti di digitalizzazione di materiale archivistico”¹². In particolare, esse specificano che «per ogni fotografia saranno prodotti tre file immagine di diverso formato: TIFF non compresso, con risoluzione di 300 dpi e profondità di colore 24 bit RGB: tali immagini costituiranno le copie master di sicurezza e serviranno inoltre per produrre tutte le successive conversioni; JPEG di alta qualità con compressione minima, con risoluzione di 150 dpi e profondità di colore 24 bit RGB, JPEG di bassa qualità con elevato fattore di compressione, con risoluzione di 72 dpi e profondità di colore 24 bit RGB, destinate al web».

Infine, le “Linea guida digitalizzazione degli archivi cartacei delle pratiche edilizie” (Ver. 1, Dicembre 2018) a cura della Comunità tematica Documenti digitali della Regione Emilia-Romagna¹³ specificano che «per quanto riguarda i formati dei file, si consiglia di scegliere sempre dei formati che siano adatti alla conservazione. Seguendo quanto indicato all’Allegato 2 del DPCM 3 dicembre 2013 Regole tecniche in materia di sistema di conservazione¹⁴, il formato migliore per i documenti testuali è il PDF/A, mentre per i formati grafici le Regole tecniche consigliano il TIFF (versione 6.0 di Adobe su cui sono costruiti gli standard ISO), ma si può utilizzare anche il PDF/A. Qualsiasi altro formato diverso da questi necessiterebbe di una verifica preventiva con ParER¹⁵ per essere certi della conservabilità».

Anche analizzando la vasta documentazione prodotta fin dagli anni ‘90 del secolo scorso dall’Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU) si individuano gli stessi formati, con alcune piccole varianti dovute per lo più all’epoca in cui tali documenti sono stati scritti. Ad esempio, le “Linee di indirizzo per i progetti di digitalizzazione del materiale fotografico” (gennaio 2004)¹⁶ dell’ICCU, con riferimento ai criteri di acquisizione delle immagini suggeriscono di «archiviare il file source nel formato TIFF così come lo restituisce lo strumento di acquisizione. Il file potrà essere successivamente trattato per la visualizzazione a monitor e per la stampa. È infatti possibile, con operazioni in automatico, ottenere dei file in formati che possono essere utilizzati in Internet. Il formato di compressione raccomandato è il JPEG». Le stesse linee di indirizzo specificano che in fase di produzione delle immagini destinate alla conservazione vanno utilizzati «formati digitali consolidati (ad esempio TIFF, PNG,

¹² Le linee guida sono disponibili all’indirizzo: <https://www.sab-campania.beniculturali.it/wp-content/uploads/2018/Digitalizzazione/Progetti-di-digitalizzazione_Linee-guida.pdf>.

¹³ Le linee guida sono disponibili all’indirizzo: <<https://digitale.regione.emilia-romagna.it/comtem/documenti/prodotti>>.

¹⁴ Tali regole tecniche sono state oggi sostituite dalle *Linee guida sulla formazione, gestione e conservazione di documenti informatici* dell’Agenzia per l’Italia Digitale e relativi allegati, disponibili all’indirizzo: <https://www.agid.gov.it/sites/default/files/repository_files/linee_guida_sul_documento_informatico.pdf>.

¹⁵ Il ParER è il polo archivistico della Regione Emilia Romagna. Cfr. <<https://poloarchivistico.regione.emilia-romagna.it>>.

¹⁶ Le linee di indirizzo sono disponibili all’indirizzo: <https://www.iccu.sbn.it/export/sites/iccu/documenti/Linee_guida_fotografie.pdf>.

PCD¹⁷)»¹⁸. Analogamente, le “Linee guida per la digitalizzazione del materiale cartografico” (maggio 2006)¹⁹ sempre dell’ICCU, dopo aver precisato che «la scelta dei formati ha un impatto fondamentale sull’accessibilità finale», propone l’adozione di formati multirisoluzione che consentono «di ingrandire l’immagine fino alla sua massima risoluzione» e costituiscono «una alternativa ormai affermata alla tradizionale strategia di produrre tre immagini diverse: copia master ad alta risoluzione, media e bassa risoluzione per gli altri possibili utilizzi. Per la cartografia di piccolo formato possono essere ancora usati i formati tradizionali come il Tiff». Per quanto riguarda gli standard tecnici da seguire nell’acquisizione e memorizzazione dell’immagine master si rinvia a «quelli individuati a livello internazionale per le collezioni digitali cartografiche: 1) formato TIFF 5 e 6; 2) risoluzione 300-400 ppi, o maggiore; 3) profondità di colore 8 bit per scala di grigi e 24 bit, o 48 in casi particolari, per carte a colori; 4) non compresso o compressione senza perdita di informazioni (LZW, PNG)». Anche le “Linee guida per la digitalizzazione di bandi, manifesti e fogli volanti” (settembre 2006)²⁰ raccomandano, per quanto riguarda la scelta dei formati di salvataggio, «di archiviare, così come lo restituisce lo strumento di acquisizione, il file source o master digitale ad alta risoluzione nel formato TIFF non compresso o eventualmente, con compressione reversibile, cioè senza perdita d’informazione (tipo lossless), da cui ottenere anche la versione Web»²¹. Per facilitare la visualizzazione in rete conviene «convertire il file master in formato JPEG riducendo il numero di ppi in modo coerente all’utilizzo previsto». Infine, le “Linee guida tecniche per i programmi di creazione di contenuti culturali digitali” (edizione italiana, 2006)²² specificano che «le immagini a matrice di punti, sia originate da operazioni di digitalizzazione, sia appartenenti alle immagini grafiche non vettoriali (come quelle create al computer, loghi, icone, disegni al tratto) dovrebbero [...] essere memorizzate utilizzando formati quali: Tagged Image File Format (TIFF), Portable Network Graphics (PNG), oppure, con attenzione alle considerazioni che seguono, anche Graphical Interchange Format (GIF), JPEG2000 o JPEG Still Picture Interchange File Format (JPEG/SPIFF)»²³.

¹⁷ Il PCD (Kodak Photo CD) era un formato proprietario sviluppato da Kodak per la memorizzazione di immagini ottenute dalla scansione delle pellicole da 35 mm mediante le unità scanner prodotte dall’azienda. Ciascun file PCD contiene una singola immagine salvata in cinque diverse risoluzioni a partire da 128x192 pixel (thumbnail), 256x384 pixel per il web, 512x764 pixel per la fruizione sul computer e sulla televisione, 1024x1536 pixel per HDTV, 2048x3072 per la stampa e 4094x6144 pixel per la stampa o l’archiviazione.

¹⁸ *Ivi*, p. 14.

¹⁹ Le linee guida sono disponibili all’indirizzo <https://www.iccu.sbn.it/export/sites/iccu/documenti/linee_guida_digit_cartografia_05_2006.pdf>.

²⁰ Le linee guida sono disponibili all’indirizzo <https://www.iccu.sbn.it/export/sites/iccu/documenti/linee_guida_bandi_sett.2006.pdf>.

²¹ *Ivi*, pag. 42.

²² Le linee guida sono disponibili all’indirizzo <<https://www.iccu.sbn.it/it/normative-standard/linee-guida-per-la-digitalizzazione-e-metadati/index.html>>.

²³ *Ivi*, p. 45.

Riguardo l'utilizzo dei formati GIF e JPEG come formati master, le linee guida fanno presente che «il formato GIF limita la codifica del colore, non consentendo di utilizzare più di 256 colorazioni diverse: è perciò generalmente utilizzato per la codifica d'immagini grafiche non vettoriali». Anche riguardo il formato JPEG occorre tenere presente «che il metodo di compressione JPEG comunemente utilizzato è del tipo lossy (con perdita di qualità)²⁴; pertanto, nella realizzazione di un master digitale, questo metodo dovrebbe essere utilizzato solo in casi particolari, quando per esempio la buona qualità delle immagini fornite dal sensore utilizzato (accertata con metodi oggettivi) consenta un margine di degrado accettabile da parte del metodo di compressione. In tutti gli altri casi deve essere privilegiato un metodo di memorizzazione reversibile (nessuna compressione o compressione senza perdita)»²⁵.

Riepilogando quanto visto fino ad ora, le varie linee guida e raccomandazioni sono concordi, da almeno venti anni, nel ritenere il TIFF come formato d'elezione per le immagini master e il JPEG come formato d'elezione per quelle derivate, salvo lievi e ben motivate eccezioni.

Le indicazioni del Piano nazionale di digitalizzazione

Anche il "Piano nazionale di digitalizzazione del patrimonio culturale" (PND), elaborato dall'Istituto centrale per la digitalizzazione del patrimonio culturale –

²⁴ I formati per immagini si distinguono in due grandi categorie: *compressi* e *non compressi*; a loro volta i formati per immagini compressi si distinguono in *lossless* (senza perdita di qualità) e *lossy* (con perdita di qualità). I primi permettono di ridurre notevolmente la dimensione dell'immagine originale, raggiungendo rapporti di compressione fino anche al 90%, ma a scapito della qualità. Si basano su algoritmi che riescono a ridurre la dimensione delle immagini eliminando quelle informazioni (ad esempio, le sfumature di colore) ed approssimando i dettagli che, secondo determinati modelli visivi, l'occhio umano fa più difficoltà a percepire. Evidentemente, più si riducono le dimensioni delle immagini, più diminuisce anche la loro qualità. La riconversione nel formato originale, partendo dai formati lossy, non permette di riottenere l'immagine originale, e la perdita di informazioni è definitiva. I formati per immagini compressi di tipo lossless, invece, cercano di ridurre la dimensione dell'immagine senza eliminare alcuna informazione; il rapporto di compressione risulta decisamente inferiore rispetto ai formati lossy, ma non c'è perdita di informazione. Se si effettua la riconversione nel formato di partenza si ottiene esattamente l'immagine originale senza alcuna perdita di qualità. Tra gli algoritmi lossless vanno menzionati: CCITT group 3 e 4 (utilizzati nella trasmissione dei fax), Flate, LZW (Lempel-Ziv-Welch), LZ77, LZ78, LZMA, RLE, Deflate (utilizzato da programmi come Winzip); tra gli algoritmi lossy va senz'altro citato l'algoritmo JPEG che si basa sulla funzione DCT (Discrete Cosine Transform). Molti formati per immagini lasciano all'utente la possibilità di scegliere se utilizzare algoritmi lossless o lossy, ma vengono normalmente utilizzati solo con una delle due tipologie. Ad esempio, il formato JPEG consente sia la compressione lossy che quella lossless, ma viene tipicamente utilizzato con la compressione lossy (in particolare, utilizzando l'algoritmo omonimo); il livello di compressione può essere scelto dall'utente: chi ha avuto modo di lavorare con le immagini JPEG, si sarà certamente imbattuto nella finestra di dialogo che chiede qual è il livello di compressione desiderato. Si noti che anche alla massima qualità concessa dalla compressione di tipo lossy, alcune informazioni vengono comunque eliminate. Tra i formati per immagini che utilizzano prevalentemente (o esclusivamente) compressione di tipo lossless si segnalano il TIFF, il PNG e il GIF.

²⁵ *Ivi*, p. 46.

Digital Library del Ministero della Cultura e pubblicato nel mese di giugno 2022²⁶, si è occupato del tema, dedicando l'intero capitolo 3 delle "Linee guida per la digitalizzazione del patrimonio culturale" all'argomento della scelta dei formati elettronici per la digitalizzazione. Le Linee guida non forniscono una trattazione esaustiva in merito a tutti i formati disponibili per la digitalizzazione del patrimonio culturale, ma si limitano ad indicare il formato e le caratteristiche generali più opportune per ciascun dominio. Innanzitutto specificano che i formati dei file prodotti in un progetto di digitalizzazione «vanno scelti tra quelli che possono maggiormente garantire l'interoperabilità tra i sistemi e la conservazione digitale, adottando, laddove possibile, formati standard internazionali e aperti. [...] Il formato per la conservazione a lungo termine²⁷ deve essere senza perdita di informazione (lossless) e non compresso, preferibilmente di tipo "grezzo" (RAW). Tra questi, per le immagini, il DNG è preferibile in quanto formato "aperto" adottato da molteplici aziende produttrici di fotocamere e software. [...] L'utilizzo di copie lossy, eventualmente compresse – come JPEG per le immagini [...] – è possibile prevalentemente per la pubblicazione on-line».

Per quanto riguarda i file immagine prodotti come output di un progetto di digitalizzazione di materiale cartaceo, le Linee guida precisano che il file master di output consiste in «un pacchetto di file comprendente il file TIFF 6.0 non compresso a 16 o 48 bit (scala di grigi o RGB), a seconda della cromia del bene» e «il file RAW non compresso, con allegato il file collaterale XMP». Specificano, inoltre, che «per la consegna dei file RAW è preferibile il formato DNG»²⁸. Relativamente alla risoluzione spaziale minima del file master, il Piano ricorda che essa «dipende da molteplici fattori quali, ad esempio, la tipologia del sensore della fotocamera o dello scanner, la dimensione del bene, ecc. Per esigenze di conservazione di posi-

²⁶ Il piano è disponibile all'indirizzo <<https://docs.italia.it/italia/icdp/icdp-pnd-docs/it/v1.0-giugno-2022/index.html>>.

²⁷ Le Linee guida rinviano, per quanto riguarda i formati adatti alla conservazione a lungo termine, all'Allegato 2 "Formati di file e riversamento" alle *Linee Guida sulla formazione, gestione e conservazione dei documenti informatici* dell'Agenzia per l'Italia Digitale (AGID).

²⁸ Le Linee guida precisano che «le specifiche tecniche dei pacchetti dei file master possono essere adattate in base a documentate esigenze specifiche del progetto di digitalizzazione. Alcuni fattori che possono condizionare la scelta del numero dei file componenti il pacchetto "master" e la tipologia del formato di ciascuno di questi file sono la tipologia dei materiali analogici da digitalizzare, le caratteristiche dei macchinari e delle tecniche di digitalizzazione più appropriate in relazione agli oggetti analogici, il numero degli oggetti da digitalizzare nell'ambito del progetto e i costi di storage connessi all'archiviazione digitale dei file master. Nel caso, ad esempio, della digitalizzazione bidimensionale di documenti cartacei, quali periodici a stampa, come giornali e quotidiani, è possibile richiedere come file master unicamente il file TIFF secondo le specifiche sopra indicate. Infatti, il considerevole numero di oggetti digitali prodotti nell'ambito del progetto di digitalizzazione di ambito bibliotecario si traduce in un maggiore costo di storage del gran numero di file RAW prodotti in fase di digitalizzazione. In ogni caso, la scelta dei componenti del pacchetto dei file master, delle loro caratteristiche e formati deve essere adeguatamente documentata e argomentata nel piano di gestione dei dati».

tivi e negativi fotografici in bianco e nero è necessario eseguire comunque la digitalizzazione in RGB a 48 bit».

Riguardo i file derivati, dopo aver fatto presente che «dipendono dalla specificità del progetto di digitalizzazione», vengono consigliati «file compressi in formato JPEG con lato lungo di almeno 3.000 pixel, con qualità di compressione non inferiore al 75%, aventi spazio colore sRGB o scala di grigi, a seconda della tipologia di scansione effettuata». Come alternativa al “classico” JPEG «può essere richiesto il formato contenitore HEIF (High Efficiency Image Format) o HEIC (High Efficiency Image Coding)». Viene segnalato anche «il formato standard FITS, definito dallo IAU FITS Working Group, Commission B2 Data and Documentation (Definition of the Flexible Image Transport System (FITS))».

I formati per la produzione di file master

Come si è potuto vedere, la quasi totalità delle linee guida sui progetti di digitalizzazione indica nel TIFF (Tagged Image File Format) il formato di riferimento per l’acquisizione dei file master. Il TIFF è un formato per immagini che è stato sviluppato a metà degli anni ‘80 dalla società Aldus Corporation nel tentativo di mettere d’accordo i fornitori di scanner su un formato comune per le immagini ottenute con il processo di scansione, al posto della moltitudine di formati proprietari allora esistenti. All’inizio, era un formato per immagini bitonale (solo due valori per ogni pixel, quindi bianco e nero), perché gli scanner dell’epoca erano in grado di acquisire solo immagini in bianco e nero, ma nel tempo si è sviluppato fino a comprendere prima le immagini in scala di grigi e poi quelle a colori.

Dopo due importanti bozze (che hanno ex post assunto la denominazione di versione 1.0 e 2.0), la prima versione ufficiale delle specifiche del formato (nota come versione 3.0) è stata pubblicata da Aldus Corporation nell’autunno del 1986. Nell’aprile 1987 fu rilasciata la versione 4.0, che conteneva per lo più miglioramenti minori. Nell’ottobre 1988 venne rilasciata la revisione 5.0, che aggiungeva il supporto per le immagini a colori e la compressione LZW. Il TIFF è un formato complesso, che definisce molti tag ma di cui, in genere, solo alcuni vengono utilizzati regolarmente. Ciò ha fatto sì che le varie implementazioni supportassero molti sottinsiemi diversi del formato²⁹. Questo problema fu affrontato nella revisione 6.0 delle specifiche TIFF, rilasciata nel giugno 1992, che introdusse la distinzione tra “TIFF baseline” (con le caratteristiche comuni a tutte le implementazioni) e le “estensioni TIFF”, opzionali. Ulteriori estensioni vennero poi definite in due supplementi alla specifica, pubblicati rispettivamente nel settembre 1995 e nel marzo 2002³⁰. TIFF supporta diverse forme di compressione senza perdita, tra cui la com-

²⁹ Questa situazione ha dato origine alla battuta secondo la quale “TIFF” sarebbe l’acronimo di “Thousands of Incompatible File Formats”. Cfr. Martin H. Trauth, *MATLAB Recipes For Earth Sciences*, Heidelberg: Springer, 2006, p. 198.

³⁰ Dal formato TIFF sono stati derivati nel tempo numerosi profili, come: il TIFF_UNC (TIFF

pressione LZW che è abitualmente utilizzata per le immagini a colori³¹ e i cui brevetti sono ormai scaduti. Il TIFF può utilizzare anche la compressione JPEG lossy, anche se viene raramente utilizzata. Il TIFF è tutt'ora un formato proprietario; infatti, la specifica del formato, inizialmente di proprietà di Aldus Corporation, è stata poi acquisita da Adobe Inc. nel 1994, che ne detiene ancora oggi i diritti di proprietà.

Nel caso in cui l'acquisizione delle immagini avvenga mediante l'utilizzo di una fotocamera digitale, viene sempre più spesso suggerito l'utilizzo del formato DNG (Digital Negative Image). Si tratta di un formato per immagini di tipo RAW sviluppato a partire dal 2004 da Adobe, partendo dal formato TIFF 6.0 di cui costituisce essenzialmente una variante. È aperto, proprietario ma esente da royalty, altamente compatibile e in continuo miglioramento. Dal momento che esistono oltre 500 tipi di formati RAW³² diversi – e ogni giorno ne vengono creati di nuovi – è altamente consigliato l'impiego del formato DNG, che rappresenta il formato standard "de-facto" per la categoria dei formati per immagini RAW, per archiviare le immagini acquisite in un progetto di digitalizzazione rispetto al formato RAW originario. Grazie all'utilizzo della compressione lossless, le immagini in formato DNG hanno una dimensione inferiore di circa il 20-30% rispetto alle corrispondenti immagini in formato RAW, pur conservando la stessa qualità dell'immagine.

Da ultimo, tra i formati per la produzione di file master merita una segnalazione il formato FITS (Flexible Image Transport System), adottato dalla Biblioteca Apostolica Vaticana³³. Si tratta di un formato creato nei primi anni '70 per risolvere

Uncompressed Bitmap); il TIFF_G4, (TIFF Bitmap with Group 4 Compression); il TIFF_LZW (TIFF Bitmap with LZW Compression); il TIFF_PYR (TIFF Piramidale); il TIFF/IT (TIFF for Image Technology); il TIFF/EP (TIFF for Digital Photography); il DNG_1_1, Adobe Digital Negative (DNG), Version 1.1; il GeoTIFF (TIFF for georeference and geocoding information). Per un approfondimento si veda la pagina relativa al formato TIFF e ai suoi profili sul sito della Library of Congress: <<https://www.loc.gov/preservation/digital/formats/fdd/fdd000022.shtml>>. Si noti che, mentre il formato TIFF non è mai stato riconosciuto come standard, alcuni dei suoi profili hanno ottenuto questo riconoscimento. Ad esempio, sono standard ISO il TIFF/EP (ISO 12234-2), il TIFF/IT (ISO 12639), il TIFF-F (RFC 2306) e il TIFF-FX (RFC 3949). Il riconoscimento come standard internazionale ISO garantisce che i prodotti siano sicuri, affidabili e di buona qualità.

³¹ Oltre alla versione non compressa, che è la più comunemente utilizzata, le immagini in formato TIFF possono anche essere compresse sia con algoritmi di compressione lossless che lossy. In particolare possono essere utilizzati i seguenti algoritmi: Packbits, un tipo di compressione lossless creata da Apple ed utilizzata per la codifica RLE (Run Length Encoding); CCITT (codifica Huffman), utilizzata in particolare per la codifica di immagini bionali (ovvero, in bianco e nero); il "Gruppo 3" e il "Gruppo 4" sono particolarmente conosciuti perché utilizzati nella trasmissione di immagini via fax; LZW (Lempel, Ziv e Welch, i cognomi dei creatori dell'algoritmo), un tipo di compressione lossless che supporta diverse profondità di bit; JPEG, una tecnica di compressione lossy molto diffusa e in grado di ridurre drasticamente le dimensioni; Deflate, una compressione lossless che utilizza le tecniche Huffman e LZ77 e supporta anche diverse profondità di bit.

³² Un file RAW è costituito dai dati di immagine non compressi e non elaborati acquisiti dai sensori di una fotocamera digitale o di uno scanner. L'immagine ottenuta presenta un elevato livello di dettaglio ma anche grandi dimensioni.

³³ Cfr. Crescenzo Tortora, *FITS: il formato per immagini e tabelle usato in astronomia e per i manoscrit-*

il problema della grande varietà di formati elettronici esistenti in quel periodo per archiviare le immagini e i dati in astronomia e astrofisica. In quell'epoca ogni istituzione aveva un proprio modo di conservare i dati raccolti, per cui si rese necessario stabilire uno standard comune allo scopo di evitare il dispendio di tempo e di risorse solitamente necessari per sviluppare un software personalizzato per convertire i dati ricevuti nel formato utilizzato nell'istituzione di provenienza a quello utilizzato dall'istituzione ricevente, e viceversa. Oggi il FITS è diventato il formato elettronico più comunemente utilizzato nel campo dell'astronomia e dell'astrofisica; tuttavia, negli ultimi anni, grazie alla sua flessibilità, è stato utilizzato anche in altri settori, come quello dei beni culturali. Ad esempio, come anticipato, la Biblioteca Apostolica Vaticana ha adottato il FITS come formato predefinito per l'archiviazione e la conservazione delle immagini del suo imponente progetto di digitalizzazione degli 80.000 manoscritti più importanti del mondo³⁴. La Biblioteca Apostolica Vaticana è stata la prima istituzione a decidere di utilizzare il formato FITS non per la conservazione delle immagini catturate dai radiotelescopi ma di quelle acquisite dalla digitalizzazione del patrimonio librario e la sua strada potrebbe essere seguita anche da altre istituzioni culturali che hanno in corso o che intendono avviare progetti di digitalizzazione³⁵. Il FITS inizia a essere utilizzato anche nel settore sanitario per archiviare e conservare i dati dei sistemi di imaging (ad esempio, la tomografia computerizzata nella medicina nucleare). Recentemente il formato FITS è stato riconosciuto come standard UNI 11845:2022 "Processi di gestione della conservazione a lungo termine di immagini digitali con l'uso del formato FITS". La norma, pubblicata nel catalogo nazionale UNI il 20 gennaio 2022, definisce caratteristiche e requisiti funzionali che un archivio basato sull'uso del formato FITS deve soddisfare per l'idoneità a lungo termine della conservazione delle immagini digitali.

ti della Biblioteca Vaticana, <<https://www.agendadigitale.eu/cultura-digitale/fits-il-formato-per-immagini-e-tabelle-usato-in-astronomia-e-per-i-manoscritti-della-biblioteca-vaticana>>.

³⁴ Cfr. Stefano Allegrezza, *Analisi del formato FITS per la conservazione a lungo termine dei manoscritti. Il caso significativo del progetto della Biblioteca Apostolica Vaticana*, «DigItalia. Rivista del digitale nei beni culturali», 6 (2011), n. 2, p. 43-72 (49), <<https://digitalia.cultura.gov.it/article/view/476>>; Luciano Ammenti, *BAV and the FITS (Flexible Image Transport System) format 40 Years of Experience in Long-Term Digital conservation*, European Week of Astronomy and Space Science (EWASS), July, 1-6, Rome (Italy), <https://www.vatlib.it/moduli/Ammenti_EWASS2012.pdf>; Paola Manoni - Ángela Núñez Gaitán - Irmgard Schuler, *The Vatican Library's Digital Preservation Project*, contributo presentato a: IFLA WLIC 2018 – Kuala Lumpur, Malaysia – Transform Libraries, Transform Societies in Session 160 – Preservation and Conservation with Information Technology, <<http://library.ifla.org/id/eprint/2113/1/160-manoni-en.pdf>>.

³⁵ Un altro fattore importante per l'adozione diffusa del FITS è la facilità di scrittura del codice per il suo utilizzo, data dal fatto che tutta la documentazione è sempre stata disponibile pubblicamente. Di conseguenza, nel corso del tempo è stata sviluppata una grande quantità di software, principalmente per la visualizzazione del materiale e la conversione da e verso il formato FITS. L'importanza di questa caratteristica non può essere sopravvalutata, perché una delle cause principali dell'obsolescenza è il legame diretto tra un formato elettronico e l'azienda che lo ha creato (il cosiddetto vendor lock-in), che comporta grossi problemi nel caso in cui l'azienda stessa dovesse smettere la sua attività lasciando gli utenti senza gli strumenti necessari per gestire quel formato.

I formati per la produzione di file derivati

Le varie linee guide sono generalmente concordi nell'indicare nel JPEG (Joint Photographic Experts Group) il formato di elezione per la produzione di immagini derivate, da destinare alla consultazione sia attraverso le postazioni messe a disposizione dell'utenza in sala studio che da remoto attraverso il web. Creato nel 1992 dal Joint Photographic Experts Group (a cui deve il nome), è il formato per immagini³⁶ oggi più famoso e utilizzato, ben documentato e formalizzato nella norma ISO/IEC 10918. Il JPEG consente di comprimere le immagini mediante una codifica sia lossy³⁷ che lossless, ma solo la versione lossy è molto utilizzata; questo significa che la sua qualità diminuisce con il diminuire della dimensione del file (ovvero, all'aumentare della compressione), ma solitamente è possibile ottenere un buon compromesso tra dimensioni e qualità. La codifica JPEG è progettata per lavorare meglio con immagini a tonalità continua, come le fotografie, mentre funziona male con le immagini line art, che presentano blocchi di colore uniformi e transizioni brusche. La rappresentazione del colore è limitata a 8 bit per canale; altri formati possono offrire una migliore risoluzione del colore. Le immagini JPEG possono contenere metadati EXIF e XMP³⁸. Il formato JPEG è la scelta migliore per le immagini derivate, ma di solito non è la scelta migliore come formato master, anche se praticamente tutti gli scanner e le fotocamere digitali oggi presenti in commercio consentono di produrre immagini in formato JPEG. Tuttavia, se si sta digitalizzando una collezione di vecchi manifesti o di documenti di grande formato, che produrrebbero immagini "pesanti" se si utilizzasse il formato TIFF, la scelta del formato JPEG potrebbe essere accettabile.

Anche per il JPEG si possono fare considerazioni simili a quelle fatte per il formato TIFF: si tratta di un formato che sta cominciando a diventare obsoleto e già vi sono numerosi candidati per la sua successione. Uno di questi è il formato JPEG2000³⁹ che, agli inizi del XXI secolo, sembrava molto promettente⁴⁰. Si tratta di un formato per immagini che è stato formalizzato come standard ISO/IEC 15444. Il JPEG2000 presenta una serie di vantaggi: gestisce bene sia la codifica con compressione lo-

³⁶ Per essere del tutto precisi, JPEG è un formato contenitore, non un formato di file. Quello che viene chiamato "file JPEG" è quasi sempre un file JFIF (JPEG File Interchange Format). Alcuni file contengono solo un flusso di dati JPEG grezzo. In genere possono essere letti, ma non hanno la "firma" di un file JFIF e quindi non sono facilmente identificabili dal software come JPEG.

³⁷ La versione con compressione lossless, nota come JPEG-LS, non è molto diffusa.

³⁸ Le immagini fornite dalle fotocamere hanno spesso già i metadati EXIF relativi al dispositivo e al timestamp, comprese le informazioni geospaziali se la fotocamera è dotata di GPS.

³⁹ La somiglianza del nome con JPEG deriva dalla loro origine comune, il Joint Photographic Experts Group, ma le somiglianze si fermano qui perché si tratta di due formati molto diversi. Come il JPEG, anche il JPEG2000 è un formato contenitore e uno standard di codifica e compressione piuttosto che un formato di file. I due principali formati di file sono JP2 e JPX. JP2 è il più semplice e il più utilizzato.

⁴⁰ Cfr. Franco Lotti, *La qualità delle immagini nei progetti di digitalizzazione*, «DigItalia. Rivista del digitale nei beni culturali», 1 (2006), n. 2, p. 22-37, <<https://digitalia.cultura.gov.it/article/view/300>>.

lossless che lossy; consente di memorizzare regioni selezionate di un'immagine con una qualità superiore rispetto ad altre; è progettato per facilitare la visualizzazione di parti di un'immagine di grandi dimensioni senza la necessità di visualizzare l'intera immagine; consente di utilizzare fino a 38 bit di colore per canale. Supporta più livelli di qualità, in modo che lo stesso file possa contenere una miniatura, un'immagine a qualità intermedia e un'immagine stampabile ad alta risoluzione. Fornisce un buon supporto per molti tipi di metadati. Nonostante tutto ciò, la mancanza di implementazioni a livello software e la scarsa adozione da parte degli istituti culturali ne mette in dubbio la sostenibilità a lungo termine, tanto che alcune delle aziende che lo avevano inizialmente sostenuto hanno poi perso interesse.

Un altro possibile successore del formato JPEG è il formato JPEG XR (eXtended Range)⁴¹. È stato sviluppato da Microsoft e nel passato è stato presentato prima con il nome di "Windows Media Photo" e successivamente con quello di "HD Photo". È un formato approvato da Joint Picture Working Group e formalizzato come standard ISO/IEC 29199-2. Le immagini in questo formato hanno generalmente l'estensione .JXR. Sono supportate sia la codifica lossy che quella lossless. Fornisce colori fino a 48 bit e si suppone che offra una compressione migliore rispetto a JPEG. Può contenere metadati XMP e profili colore ICC. Purtroppo anche questo formato non ha riscosso un grande successo; in particolare non è molto adottato al di fuori del mondo Microsoft⁴². Microsoft detiene i brevetti di JPEG XR, ma ha rilasciato la specifica sotto la promessa di Microsoft Open Specification, che consente l'implementazione senza dover pagare una licenza.

Rimanendo nel mondo dei formati riconducibili, in qualche modo, al Joint Photographic Experts Group, si segnala che sono in corso iniziative per la standardizzazione sia di formati per immagini, come il JPEG XL (eXtended Long-term)⁴³ – un formato per immagini di tipo raster, nato nel 2017, che supporta sia la compressione lossy che lossless ed è progettato per ottenere una compressione più efficiente dei formati preesistenti e arrivare a sostituirli in tutte le situazioni – che di algoritmi di compressione delle immagini, come il JBIG2, – uno standard di compressione delle immagini a due livelli (immagini bianco e nero), sviluppato dal Joint Bi-level Image Experts Group e adatto sia per la compressione senza perdita che per quella con perdita⁴⁴. Al momento in cui si scrive non è possibile avanzare

⁴¹ Si noti che, nonostante il nome, non è correlato in alcun modo ai formati JPEG o JPEG2000.

⁴² Ad oggi, né la codifica JPEG XR né il formato di file JXR sono ampiamente adottati, sebbene molte applicazioni Microsoft possano leggere o scrivere file JXR, come Microsoft Foto in Windows 10 o le applicazioni di Microsoft Office, come Word e PowerPoint.

⁴³ La "L" che sta per "Long-term" è stata inclusa per riflettere l'intenzione degli autori di creare un formato capace di sostituire il precedente JPEG e durare altrettanto a lungo.

⁴⁴ Questo tipo di compressione riesce a creare flussi di dati da 3 a 5 volte più piccoli rispetto al formato Fax Group 4 e da 2 a 4 volte più piccolo del JBIG (altresì chiamato JBIG1). Esso può essere utilizzato, ad esempio, per produrre file PDF più piccoli dove si usano loghi in bianco e nero o testi acquisiti in formato immagine.

ipotesi sulla loro adozione futura, ma si tratta comunque di formati che meritano una certa attenzione.

Accanto ai formati che hanno una qualche relazione con il JPEG, ve ne sono altri che sono stati sviluppati da gruppi di lavoro o aziende del tutto diversi. Uno di questi è il WebP, un formato relativamente nuovo nel panorama delle immagini, ma che sta rapidamente aumentando la sua diffusione grazie al fatto che viene sempre più spesso utilizzato come formato per le immagini pubblicate sul web. Il formato, che è stato sviluppato da Google insieme con il WebM (il formato per i contenuti audiovisivi), è offerto sotto la licenza open-source BSD, per cui è libero da vincoli di brevetto. Può utilizzare sia una compressione lossless, basata sull'impiego dell'algoritmo video VP8, che lossy, basata sull'algoritmo LZ77. Come il formato PNG consente l'utilizzo della trasparenza (attraverso il canale alfa a 8 bit) e la possibilità di creare immagini animate attraverso la memorizzazione di sequenze di immagini, come nel formato GIF. Può contenere metadati EXIF e XMP ed incorporare profili di colore ICC. La qualità risultante delle immagini è superiore rispetto a quella del formato JPEG.

Un altro formato per immagini che si sta diffondendo molto rapidamente è l'HEIC (High Efficiency Image Container), una variante del formato HEIF (High Efficiency Image Format), tradizionalmente utilizzato da Apple nei suoi dispositivi mobili. L'HEIF non è semplicemente un formato per immagini, bensì un formato contenitore pensato per ospitare oggetti digitali di varia natura, come singole immagini, "raffiche" (burst) di scatti, intere gallerie di immagini, scatti multi-esposizione per la fotografia HDR e informazioni sulla profondità di campo. Quando questi oggetti vengono compressi utilizzando l'algoritmo di compressione HEVC (High Efficiency Video Coding) – meglio noto come H.265 – l'intera raccolta ha una dimensione che è circa pari alla metà rispetto ad altri formati di file immagine, come il JPEG; inoltre assume la forma del formato HEIC (con estensione .heic o .heics, a seconda del numero di immagini contenute)⁴⁵. L'HEIC salva le immagini con una qualità superiore a quella del JPEG, utilizzando al contempo meno spazio grazie a una tecnologia di compressione avanzata. Un file HEIC contiene anche i metadati che descrivono le dimensioni, la risoluzione, la posizione e altro ancora di ciascuna immagine.

Anche il formato PNG (Portable Network Graphics) è stato indicato come possibile successore del formato JPEG. Sviluppato tra il 1996 e il 1998 per aggirare le limitazioni e i problemi di brevetto del formato GIF (vedi *infra*)⁴⁶, è stato creato appo-

⁴⁵ Per rendere più o meno efficiente la memorizzazione di file multimediali all'interno del contenitore HEIF è possibile scegliere fra diversi profili di compressione che vanno dall'8 bit al 16 bit lossless (compressione impercettibile). Apple ha scelto di utilizzare un profilo HEVC a 10 bit con un livello di compressione media per garantire una qualità dell'immagine comparabile a quella del JPEG – che invece limita la profondità del colore a 8 bit – pur occupando circa la metà dello spazio.

⁴⁶ Il formato GIF presentava allora diversi inconvenienti: oltre a richiedere una licenza per il brevetto, consentiva una profondità di colore di soli 8 bit (quindi immagini con al massimo 256 colori) e non era, pertanto, adatto per immagini di qualità fotografica.

sitamente privo di brevetto e con una palette di colori molto più ampia, arrivando ad una profondità di colore di 16 bpp per le immagini in scala di grigi e di 48 bpp per le immagini truecolor (16 bit per canale), oltre ad un canale alfa per la trasparenza. A differenza di JPEG, che si basa sulla compressione DCT, PNG utilizza una compressione di tipo lossless grazie all'algoritmo Deflate, una variante dell'algoritmo LZW simile a quello utilizzata dai formati GIF e TIFF. Il vantaggio principale di PNG rispetto a JPEG è che la compressione è senza perdita, il che significa che non c'è perdita di qualità ogni volta che viene aperto e salvato di nuovo. Il PNG gestisce bene anche le immagini dettagliate e ad alto contrasto. Per questo motivo il PNG è spesso il formato di file predefinito per la cattura delle schermate, in quanto è in grado di fornire una rappresentazione quasi perfetta pixel per pixel dello schermo, anziché comprimere gruppi di pixel. È un tipo di formato per immagini particolarmente apprezzato dai web designer, perché, a differenza del JPEG, è in grado di gestire immagini con sfondi trasparenti o semitrasparenti; per questo è molto utilizzato per produrre elementi grafici da utilizzare sul web⁴⁷ (come loghi, banner, infografiche) ed è in grado di offrire una leggibilità del testo molto migliore rispetto al JPEG; questo aspetto lo rende un'ottima scelta per l'acquisizione di materiale che comprende sia immagini che testo. A differenza del formato GIF, il PNG è un formato a immagine singola e non supporta le animazioni; tuttavia, ne esiste una variante chiamata MNG che è utilizzata per creare immagini animate, ma poche applicazioni supportano questa funzione. Il formato è stato riconosciuto standard ISO/IEC 15948.

Il formato PDF (Portable Document Format) non è in realtà un formato di file immagine (per la precisione è un formato di descrizione di pagina) ma si tratta comunque di uno dei formati più comuni per l'acquisizione di documenti cartacei. Adobe ha progettato il formato in modo che i metadati, ovvero le informazioni descrittive sul file, possano essere acquisiti da qualsiasi applicazione. Inoltre, la maggior parte degli utenti di computer ha familiarità con il formato e con il suo utilizzo. I PDF dispongono di una tecnologia di riconoscimento ottico dei caratteri (OCR) integrata, il che significa che il testo può essere ricercato. Ormai quasi tutte le apparecchiature di scansione – siano essi scanner piani, scanner planetari, scanner documentali od altro – sono in grado di acquisire scansioni nel formato PDF nativo (ovvero prodotto direttamente dalla macchina).

Vi sono poi altri formati che, pur essendo in generale ampiamente utilizzati, non sono utilizzabili in progetti di digitalizzazione né come file master né come file derivati. Ad esempio, il formato GIF (Graphics Interface Format), è un formato di file immagine leggero e senza perdite, ideale per la grafica di piccole dimensioni sui siti web e nelle e-mail. Supporta elementi trasparenti ed è utilizzato per le imma-

⁴⁷ Le immagini in formato PNG sono fruibili con tutti i principali browser (Chrome, Edge, Firefox, Internet Explorer, Opera, Safari) e con tutti i principali sistemi operativi ed i loro editor di immagini standard.

gini animate (cd. "GIF animate", molto utilizzate sul web). Tuttavia, le immagini GIF sono limitate a 256 colori simultanei, quindi non possono raggiungere la qualità fotografica. Infine, va citato il BMP (Bitmap), un vecchio formato introdotto nel 1990 con il sistema operativo Windows 3.0 e utilizzato quasi esclusivamente nei sistemi Windows. Con questo formato le immagini non vengono compresse e, rispetto ad altri formati compressi, occupa una maggior quantità di spazio sui dispositivi di archiviazione. È stato per lo più sostituito dal TIFF (anche se esistono ancora oggi degli scanner che acquisiscono immagini nel formato BMP, più che altro per ragioni di compatibilità).

TIFF e JPEG sono ancora una buona scelta?

Il TIFF è considerato il formato di elezione per la produzione di immagini master nei progetti di digitalizzazione. Non c'è dubbio che negli anni '80 e '90 del secolo scorso, quando la potenza di calcolo e la memoria degli elaborati erano limitati, il TIFF costituisse la miglior scelta possibile. Ma possiamo dire che lo sia ancora oggi? Purtroppo la risposta non sembra così scontata. Infatti, il TIFF presenta diversi vantaggi e non sempre rappresenta la soluzione migliore⁴⁸. Innanzitutto, è un formato datato: come si è visto, l'ultima specifica è stata la 6.0 che risale al 1992⁴⁹. In secondo luogo, è uno standard "de-facto", ma non uno standard "de-jure". Ciò significa che è diventato uno standard grazie alla sua ampia adozione, ma in quasi quaranta anni non ha ottenuto alcun riconoscimento ufficiale da parte di un organismo di standardizzazione (sebbene alcuni dei suoi profili lo abbiano ricevuto⁵⁰). Inoltre, le specifiche sono ambigue su alcuni punti ed esistono molte varianti del formato TIFF che non funzionano tutte allo stesso modo perché si sono evolute autonomamente senza aderire ad alcuno standard. Di conseguenza, manca un riferimento univoco e per avere un quadro completo è necessario consultare anche le diverse varianti (di solito pubblicate come "note" nei siti delle aziende che hanno sviluppato le varianti). In terzo luogo, il TIFF consente l'utilizzo di tag standard per contenere i metadati di base dell'immagine (come i tag riservati per i metadati EXIF e XMP), ma permette anche di inserire tag personalizzati, e questo potrebbe costituire un problema di interoperabilità perché tali tag potrebbero non essere riconosciuti dai vari software.

⁴⁸ Cfr. Thomas Zellmann, *Is TIFF good enough?*, March 14, 2017, <<https://www.foxit.com/blog/is-tiff-good-enough>>.

⁴⁹ In linea generale, il fatto che un formato elettronico non venga più aggiornato non rappresenta necessariamente un fattore negativo ai fini della conservazione a lungo termine, anche se effettivamente tre decenni rappresentano "un'eternità" dal punto di vista dell'obsolescenza tecnologica. Ad esempio, il formato TXT (plain text o "puro testo") è rimasto sostanzialmente lo stesso da quasi sessant'anni, essendo basato sulla elementare codifica ASCII del testo; ciò lo rende non solo uno dei formati più trasparenti oggi esistenti, ma anche un formato con ottime caratteristiche di conservabilità a lungo termine, tant'è che numerosi standard e raccomandazioni (ad esempio i "Request for Comments" (RFC) del W3C) vengono ancora oggi prodotti in un formato di puro testo.

⁵⁰ Cfr. nota 30.

Occorre poi considerare che il TIFF non offre molte possibilità di compressione: di solito, quando le immagini sono compresse, viene utilizzata la compressione LZW, che è di tipo lossless e funziona bene per le immagini al tratto ma meno bene per le immagini a colori. Poiché le immagini TIFF compresse con questo algoritmo richiedono una maggiore potenza di elaborazione per l'apertura e la chiusura, ciò comporta un aumento dei tempi, ad esempio, quando si accede alle immagini presenti in un archivio di grandi dimensioni. Inoltre, un altro limite del TIFF è che la dimensione dei file immagine non può superare i 4 gigabyte⁵¹.

Ancora: il TIFF non è di per sé ricercabile. È stato creato espressamente per memorizzare le immagini, ma non il testo. Per rendere il TIFF ricercabile, si possono utilizzare dei tag privati, ma tali tag possono funzionare all'interno di una istituzione culturale e non è detto che funzionino anche all'esterno, ad esempio quando si scambiano immagini con altre istituzioni.

Infine, dal punto di vista della fruibilità, le immagini TIFF possono essere visualizzate utilizzando alcune applicazioni predefinite dei vari sistemi operativi (ad esempio, il "Visualizzatore foto" o l'app "Foto" di Windows 10) ma non attraverso i vari browser (per i quali è necessario installare delle apposite estensioni) che costituiscono lo strumento principale attraverso cui avviene la fruizione delle immagini da parte del pubblico.

Analoghe riflessioni potrebbero essere fatte per il formato JPEG: possiamo dare per scontato che oggi sia la scelta migliore come formato derivato? Come si è visto nei paragrafi precedenti, ci sono molti formati che hanno prestazioni migliori rispetto al JPEG, ovvero formati che, a parità di qualità, consentono di produrre immagini di dimensioni inferiori (oppure, a parità di dimensioni delle immagini, consentono una qualità superiore). È in corso una vera "battaglia" tra i "contendenti" alla successione del JPEG ma quello che è certo è che il JPEG non riuscirà a mantenere a lungo la sua supremazia e sarà, prima o poi, "rimpiazzato" da altri formati per immagini.

Il nuovo formato PDF/R (PDF/Raster)

Recentemente è stato proposto un nuovo formato che sembra promettere di poter diventare il formato di elezione per l'acquisizione delle immagini in molti progetti di digitalizzazione. Si tratta del PDF/R (la "R" sta per "Raster")⁵², un nuovo profilo del formato PDF – che va ad aggiungersi a quelli già ampiamente noti, come il PDF/A, il PDF/X, il PDF/E, il PDF/UA ecc. – specificatamente pensato per l'archiviazione e il trasporto di immagini raster sia a pagina singola

⁵¹ Esiste una variante chiamata BigTIFF che supera questo limite ma non è compatibile con TIFF, anche se le immagini TIFF potrebbero essere convertite nel formato BigTIFF (diventando un po' più grandi) senza alcuna perdita semantica.

⁵² Per ulteriori informazioni si faccia riferimento al sito <<https://pdfraster.org/specification>>.

che multi pagina, con particolare riferimento a quelle acquisite nei progetti di digitalizzazione del patrimonio culturale archivistico e librario. La nascita del PDF/R risale a qualche anno fa, quando il TWAIN Working Group⁵³ ne sviluppò la specifica originale chiamandola con il nome di “PDF/Raster”⁵⁴. Si tratta di una specifica documentata pubblicamente e sviluppata grazie alla collaborazione tra due associazioni internazionali che sviluppano e promuovono standard, il TWAIN Working Group, che ha dato origine al concetto di PDF/raster, e la PDF Association⁵⁵, che ha messo a disposizione le competenze e le conoscenze della tecnologia PDF, nonché i contatti con le aziende produttrici di strumenti software per il formato PDF al fine di garantire la presenza di punti di vista diversi. A metà del 2017, la specifica è stata presentata al sottocomitato dell’ISO/TC171/SC2⁵⁶ per chiederne la standardizzazione; quest’ultima è arrivata dapprima con l’approvazione nel mese di aprile 2020 e poi con la pubblicazione, nel mese di luglio 2020, dello standard ISO 23504-1:2020 “Document management applications — Raster image transport and storage — Part 1: Use of ISO 32000 (PDF/R-1)” che sancisce il riconoscimento del formato come standard. Una delle modifiche introdotte durante il processo di standardizzazione in seno all’ISO, è stata l’adozione del nome “PDF/R” al posto di “PDF/Raster”. Il PDF/R, essendo un sottoinsieme ristretto del formato PDF, è molto più semplice da generare e interpretare rispetto a quest’ultimo, consentendo, come vedremo, di sostituire non solo il formato TIFF per l’acquisizione e la conservazione ma anche il JPEG per la fruizione delle immagini acquisite. Essendo un profilo del formato PDF, il PDF/R è a tutti gli effetti un valido PDF, quindi si integra perfettamente in tutti i flussi di lavoro che si basano sull’utilizzo di questo formato. Inoltre, non è prevista alcuna estensione particolare per identificare un’immagine in formato PDF/R: l’estensione è .pdf, la stessa utilizzata dal

⁵³ TWAIN è un protocollo standard aperto e gratuito che facilita la comunicazione tra applicazioni e dispositivi di acquisizione di immagini. Il TWAIN working group è responsabile del mantenimento e del miglioramento dello standard TWAIN nell’ambito del progetto TWAIN Direct, per collegare gli scanner alle applicazioni attraverso le reti. Il gruppo di lavoro TWAIN si descrive come un’organizzazione senza scopo di lucro con l’unico obiettivo di promuovere uno standard pubblico universale che colleghi applicazioni e dispositivi di acquisizione delle immagini. Cfr. <<https://twain.org>>.

⁵⁴ La specifica PDF/raster presentata a ISO/TC 171/SC2 è disponibile presso il TWAIN Working Group o la PDF Association. Cfr. la pagina “PDF/Raster 1.0” sul sito della PDF/A Association <<https://www.pdfa.org/resource/pdfrafter-1-0/>>.

⁵⁵ La PDF Association ha sede a Berlino, in Germania, e gestisce un’organizzazione no-profit interamente controllata, la PDF Association, Inc. negli Stati Uniti. La sua missione è «to promote Open Standards-based electronic document implementations using PDF technology through education, expertise and shared experience for stakeholders worldwide». Cfr. <<https://www.pdfa.org/about-us>>.

⁵⁶ Cfr. la pagina del catalogo ISO relativa allo standard: <<https://www.iso.org/standard/75804.html>>. Il gruppo di lavoro che se ne è occupato è il Working Group 8 (WG8) all’interno del sottocomitato 2 del comitato tecnico 171, denominato ISO/TC 171/SC 2 *Document file formats, EDMS systems and authenticity of information*.

formato PDF e da tutti gli altri profili⁵⁷. Questo costituisce un enorme vantaggio rispetto al TIFF, perché è possibile “aprire” una immagine in formato PDF/R utilizzando qualsiasi lettore PDF (semplicemente facendo doppio click sulla sua icona) o qualsiasi browser.

Le specifiche del formato prevedono due varianti del PDF/R: la prima, conforme allo standard ISO 32000-1 (PDF 1.7), non ammette la cifratura; la seconda, conforme allo standard ISO 32000-2 (PDF 2.0), consente invece la cifratura. La prima variante può essere visualizzata con qualsiasi visualizzatore compatibile con il formato PDF 1.7; la seconda può essere visualizzata con qualsiasi programma in grado di gestire il formato PDF 2.0.

Il PDF/R presenta numerosi vantaggi rispetto ad altri formati comunemente utilizzati nei progetti di digitalizzazione, come il TIFF e il JPEG. Innanzitutto va fatto osservare che, al contrario del TIFF che si basa su specifiche non standard e su estensioni proprietarie, il PDF/R è stato riconosciuto standard ISO 23504-1, che si basa a sua volta sullo standard ISO 32000 (formato PDF). Essendo stato riconosciuto come standard ISO, non presenta restrizioni di alcun tipo sotto il profilo della proprietà intellettuale.

Dal punto di vista della qualità delle immagini, consente di produrre immagini bitonali, in scala di grigi e a colori; la profondità di colore per le immagini a colori o in scala di grigi può essere di 8 o 16 bit per canale. È possibile specificare un profilo colore ICC: sebbene quello consigliato sia lo spazio colore sRGB, è possibile utilizzare altri spazi colore. Non esiste un limite esplicito alla risoluzione spaziale di un'immagine.

A differenza del formato JPEG, il PDF/R consente di inserire dei metadati XMP basati sull'XML.

Altra caratteristica interessante del formato PDF/R è la sua versatilità: al pari del TIFF e al contrario del JPEG, che è limitato a contenere una sola immagine, le im-



⁵⁷ L'identificazione del formato avviene tramite la presenza di una ben precisa sequenza di *magic number* all'interno della sequenza di bit che costituisce una immagine in formato PDF/R. Si tratta, in particolare, di una riga di commento nel *trailer* PDF, situata immediatamente prima della riga contenente "startxref". La forma del commento è "%PDF-raster-x.y", dove x è il numero di versione maggiore e y il numero di versione minore. Ad esempio, per una immagine conforme al profilo PDF/R-1, ovvero alla specifica iniziale (ISO 23504-1:2020), il commento assume la forma "%PDF-raster-1.0".

magini in formato PDF/R possono contenere la scansione di più pagine (PDF/R multi pagina); inoltre, possono essere presenti contemporaneamente pagine a colori, in scala di grigio e in bianco e nero, così da ottenere la massima compressione ed ottimizzare le dimensioni. Questa caratteristica, che non trova riscontro in nessun altro formato per immagini, è estremamente utile. Si pensi, per fare un esempio, ad un progetto di digitalizzazione dell'archivio delle pratiche edilizie di un comune: con il PDF/R è possibile scansionare tutti i documenti contenuti all'interno di una pratica generando un unico file dove la documentazione amministrativa è acquisita in bianco e nero e gli elaborati tecnici sono acquisiti a colori, consentendo quindi la massima ottimizzazione delle dimensioni. In aggiunta, il PDF/R consente di dividere le pagine in "strisce"; questa funzione risulta estremamente utile, ad esempio, per facilitare la scansione di documenti di grande formato (come gli elaborati tecnici) o ad altissima risoluzione.

Il PDF/R consente anche la crittografia con password e certificati digitali e può contenere firme digitali anche multiple, e questo potrebbe certamente tornare molto utile, ad esempio nei progetti di digitalizzazione massiva di documenti con la certificazione di processo prevista dall'Allegato 3 alle "Linee guida AGiD sulla formazione, gestione e conservazione del documento informatico"⁵⁸.

È possibile generare immagini in formato PDF/R sia senza compressione che con compressione (utilizzando diversi algoritmi di compressione, compreso JPEG⁵⁹ e fax CCITT Group 4 con compressione di tipo lossless). Questo significa che nella generalità delle situazioni può essere utilizzato non solo come formato master al posto del TIFF (ad esempio, senza compressione o con compressione fax CCITT Group 4), ma anche come formato derivato al posto del JPEG (utilizzando, appunto, la compressione JPEG). Si noti che, quando un'immagine in formato PDF/R viene generata senza compressione, ha una qualità equivalente a quella del TIFF non compresso; se, invece, si utilizza la compressione JPEG, la qualità è equivalente all'analoga immagine in formato JPEG. Le immagini possono essere create direttamente attraverso sistemi con prestazioni molto ridotte come gli scanner e le periferiche multifunzione⁶⁰, essendo il PDF/R più semplice del TIFF e con meno varianti rispetto a quest'ultimo.

Il vantaggio principale del formato PDF/R è che le immagini acquisite in formato PDF/R possono essere visualizzate, modificate e stampate con qualsiasi programma capace di gestire i file PDF, trattandosi di un profilo del formato PDF. Ad esempio, per la visualizzazione è possibile utilizzare un qualsiasi visualizzatore di

⁵⁸ Sia le linee guida che gli allegati sono disponibili all'indirizzo https://trasparenza.agid.gov.it/archivio19_regolamenti_0_5385.html.

⁵⁹ Le successive versioni del formato PDF/R consentiranno anche la compressione JPEG2000 e JBIG2.

⁶⁰ Sebbene al momento non vi siano strumenti che acquisiscono direttamente in formato PDF/R, è presumibile che nel breve periodo le aziende produttrici introdurranno questa funzione nelle loro apparecchiature di scansione.

documenti PDF, da Adobe Reader a Foxit Reader, compreso qualsiasi browser web. Si tratta di programmi presenti su qualsiasi dispositivo, dai computer ai tablet o agli smartphone e questo rende la fruizione delle immagini – ad esempio, quelle acquisite in un progetto di digitalizzazione – estremamente semplice, al contrario di altri formati per immagini, come il TIFF, per il quale occorre un visualizzatore apposito, non sempre presente in tutte le piattaforme o in tutti i browser, e, comunque, non sempre così semplice da usare. A prima vista questo potrebbe sembrare un vantaggio di poco conto, ma non è così. La sfida della conservazione digitale richiede di andare nella direzione di una riduzione del numero di formati elettronici da gestire e conservare (si stima che i formati elettronici oggi esistenti al mondo siano oltre ventimila) e quindi è auspicabile l'impiego di un unico formato anche nel settore della digitalizzazione dei beni archivistici e librari. Non c'è dubbio che il passaggio al formato PDF/R non potrà portare che benefici per quanto riguarda l'aspetto della conservazione a lungo termine. L'attenzione verso questo aspetto è testimoniata anche dal fatto che, come indicato nell'Allegato A della specifica del PDF/R-1, è possibile creare file PDF/R conformi anche al profilo PDF/A⁶¹ e quindi già predisposti per la conservazione a lungo termine⁶².

Conclusioni

I paragrafi precedenti mostrano come il mondo dei formati per la digitalizzazione del patrimonio archivistico e librario sia estremamente complesso e variegato, e sia oggettivamente difficile districarsi tra un formato e l'altro o individuare quello corretto per gli scopi di uno specifico progetto di digitalizzazione. Il PDF/R potrebbe venire in soccorso proponendosi come il formato universale per l'acquisizione delle immagini. Le sue caratteristiche consentono di utilizzarlo sia come formato master che come formato derivato e questo ha conseguenze davvero interessanti. Ad esempio, in un progetto di digitalizzazione si potrebbe decidere di utilizzare il PDF/R non compresso per finalità di conservazione a lungo termine e lo stesso formato, con compressione di tipo JPEG, per finalità di fruizione, andando a sostituire i formati TIFF e JPEG che, come si è visto, sono i formati at-

⁶¹ Il PDF/A è il profilo del formato PDF specificatamente creato per la conservazione a lungo termine dei documenti digitali. Ad oggi ne sono state definite quattro versioni: il PDF/A-1 (ISO 19005-1:2005 "Document management — Electronic document file format for long-term preservation — Part 1: Use of PDF 1.4 (PDF/A-1)"); il PDF/A-2 (ISO 19005-2:2011 "Document management — Electronic document file format for long-term preservation — Part 2: Use of ISO 32000-1 (PDF/A-2)"); il PDF/A-3 (ISO 19005-3:2012 "Document management — Electronic document file format for long-term preservation — Part 3: Use of ISO 32000-1 with support for embedded files (PDF/A-3)") e il PDF/A-4 (ISO/19005-4:2020 "Document management — Electronic document file format for long-term preservation — Part 4: Use of ISO 32000-2 (PDF/A-4)").

⁶² Tuttavia, l'allegato afferma che gli unici profili PDF/A cui un PDF/R può essere conforme sono PDF/A-1b e PDF/A-2b; gli altri profili di PDF/A-1 e PDF/A-2 richiedono che il testo eventualmente presente nell'immagine sia ricercabile.

tualmente più utilizzati. Per di più, in alcuni casi il PDF/R potrebbe addirittura eliminare la necessità di creare le due versioni: si pensi, ad esempio, a progetti di digitalizzazione di documentazione amministrativa prodotta nel secolo scorso, dove non sempre è necessario produrre le due versioni ma potrebbe essere sufficiente produrre una sola, magari in alta qualità; in tal caso si potrebbe proporre l'adozione del PDF/R non compresso, da utilizzare sia come formato master per finalità di conservazione che come formato derivato per finalità di fruizione. Basterebbe, quindi, un solo formato, e ciò si tradurrebbe in evidenti risparmi sia in termini di tempo che economici.

In conclusione, il PDF/R, con le sue eccellenti proprietà, aspira a svolgere il ruolo di formato per l'acquisizione, la fruizione e la conservazione di immagini acquisite nei progetti di digitalizzazione, in sostituzione dei formati elettronici, come il TIFF e il JPEG, che vengono attualmente utilizzati.

This paper aims at providing some reflections on the choice of file formats to use for imaging in archives and libraries digitisation project.

To date, this choice almost always falls on the TIFF format in the case of master images produced by scanning (or possibly on one of the RAW formats in the case of acquisition by camera) and on the JPG format in the case of derived images.

However, the scientific community wonders whether these choices are still right and whether there are other file formats that, on the one hand, give better guarantees with regard to long-term preservation and, on the other hand, allow better usability. In this regard, the adoption of the new PDF/R format, which recently gained recognition as an ISO standard and which is proposed as a replacement for the TIFF and JPEG formats, deserves careful consideration. Being a profile of the PDF format, it has very good usability features; moreover, it also has excellent long-term archiving properties, thanks to its compliance with the PDF/A specification. In this paper we will first provide an overview of the indications provided by the main national guidelines, then we will analyse the alternatives to TIFF and JPEG as file formats for the creation of master and derived files, respectively; finally, we will analyse the new PDF/R format, highlighting the features that make it the preferred file format for digitization projects in archives and libraries.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2022.