

ARCHIVI & COMPUTER

AUTOMAZIONE E BENI CULTURALI

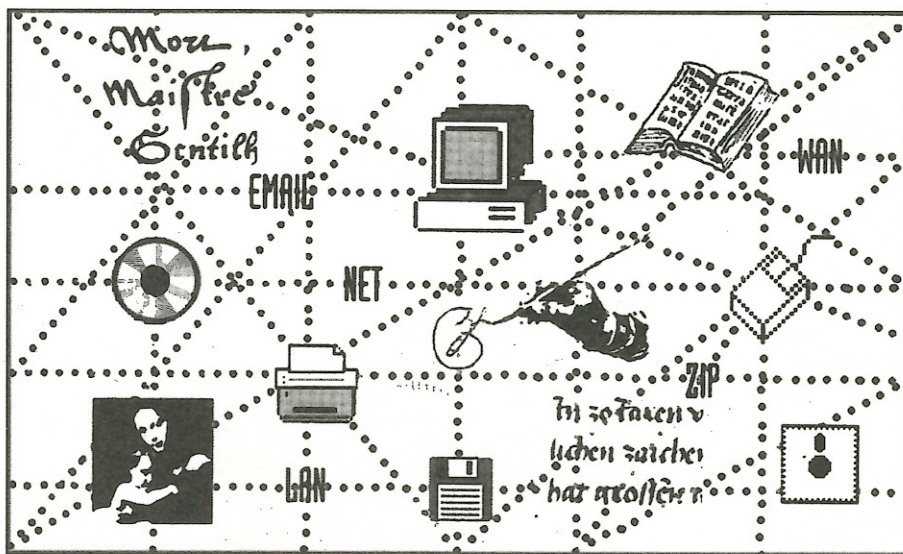
SAGGI

Stefano ALLEGREZZA, *Un nuovo standard per la conservazione digitale: il formato PDF/A-2 e le principali novità rispetto al PDF/A-1*

Valeria CALABRESE, *Gli archivi di impresa nel Biellese: dal censimento delle fonti al Portale degli archivi del tessile e della moda*

Lois ENNS, *E-Government at the City of Surrey*

Maria GUERCIO, *Custodia archivistica, ubiquità digitale*



Semestrale - Spedizione in A.P. - Filiale di Pisa - comma 34, art. 2, L. 549/95 - Pisa

Anno XXI


Titivillus

Fascicolo 2/011

HANNO COLLABORATO A QUESTO NUMERO:

Stefano ALLEGREZZA, Università degli studi di Udine

Maria GUERCIO, Università degli studi di Urbino

Valeria CALABRESE, Comune di Torino

Lois ENNS, City of Surrey, British Columbia

La redazione invita tutti coloro che hanno informazioni, opinioni, domande da porre su temi che riguardano l'automazione degli archivi a segnalarle alla Segreteria di Archivi & Computer, presso l'Archivio Storico Comunale di San Miniato (Loggiati di San Domenico, 3 - 56028 San Miniato - Pisa; tel. 0571 418381; telefax 0571 406233; e-mail archilab@comune.san-miniato.pi.it).



ARCHIVI & COMPUTER

AUTOMAZIONE E BENI CULTURALI

Anno XXI

Fascicolo 2/011

INDICE

SAGGI

- | | | |
|--------------------|---|------|
| Stefano ALLEGREZZA | <i>Un nuovo standard per la
conservazione digitale: il formato
PDF/A-2 e le principali novità
rispetto al PDF/A-1</i> | p. 5 |
| Valeria CALABRESE | <i>Gli archivi di impresa nel Biellese:
dal censimento delle fonti al Portale
degli archivi del tessile e della moda</i> | 23 |
| Lois ENNS | <i>E-Government at the City of Surrey</i> | 80 |
| Maria GUERCIO | <i>Custodia archivistica, ubiquità
digitale</i> | 92 |

Stefano ALLEGREZZA

Un nuovo standard per la conservazione digitale: il formato PDF/A-2 e le principali novità rispetto al PDF/A-1

Abstract: *The article analyses in details the second part of the standard PDF/A recently approved as ISO 19005-2:2011 Document management – Electronic document file format for long-term preservation – Part 2: Use of ISO 32000-1 (PDF/A-2). A special attention is dedicated to the reasons for its development, the level of compliance and the backward compatibility. As happened in the part 1, the criteria at the basis of this second part are connected to the acceptance of all the characteristics of the release 1.7 of the format which are not in contrast with the long-term preservation of electronic records.*

1.1.1 Introduzione

Il PDF/A, il formato¹ creato appositamente per la conservazione a lungo termine dei documenti elettronici, è ormai diventato uno standard non solo *de jure* ma anche *de facto* in molti settori ed in particolar modo in quelli legati alla gestione dei documenti elettronici e alla loro conservazione a lungo termine.

Il formato è definito dalla norma ISO 19005, una norma pensata in maniera modulare: è costituita da più parti² che sono indipendenti tra di loro e possono essere singolarmente sviluppate per seguire l'evoluzione nel tempo delle varie versioni del formato PDF a cui il PDF/A fa riferimento.

La prima parte dello standard, quella che ha definito le specifiche del PDF/A-1, è stata pubblicata dall'International Standards Organization (ISO) il 1° ottobre 2005 con la denominazione di ISO 19005-1:2005 *Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1)* e si basa sul formato PDF nella versione 1.4 (PDF Reference Version 1.4)³.

¹ Come sarà chiaro più avanti (si veda la nota 7), più che di formato di *file* si dovrebbe parlare di “profilo” (o “*subset*”) del formato PDF. Tuttavia, per semplicità, nel seguito utilizzeremo la denominazione “formato”, più usuale anche se non del tutto formalmente corretta.

² Si noti che la stessa filosofia è stata seguita nella definizione dello standard ISO che descrive le specifiche del formato PDF; la prima parte, denominata ISO32000-1:2008, descrive le specifiche del formato PDF nella versione 1.7. Seguiranno nel tempo le altre parti della norma.

³ Il *PDF Reference* è il documento che contiene le specifiche del formato PDF. Il *PDF Reference 1.4*

Il 20 giugno 2011⁴, a distanza di quasi sei anni dal rilascio della prima parte, è stata pubblicata la seconda parte dello standard con la denominazione di ISO 19005-2:2011 *Document management – Electronic document file format for long-term preservation – Part 2: Use of ISO 32000-1 (PDF/A-2)*. Questa seconda parte dello standard definisce le specifiche del formato PDF/A-2 basandosi sul formato PDF nella versione 1.7 (PDF Reference Version 1.7).

Al momento sono state pubblicate solo le prime due parti dello standard⁵, ma una terza parte, quella che definirà il PDF/A-3, si trova attualmente in fase di sviluppo⁶ con la denominazione di *ISO/DIS 19005-3 “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 la sua pubblicazione è prevista per il 2012. Nel tempo si succederanno anche le altre parti dello standard che faranno riferimento alle successive versioni del formato PDF.

La Figura 1 riassume sinteticamente le denominazioni delle parti che attualmente compongono lo standard ISO 19005.

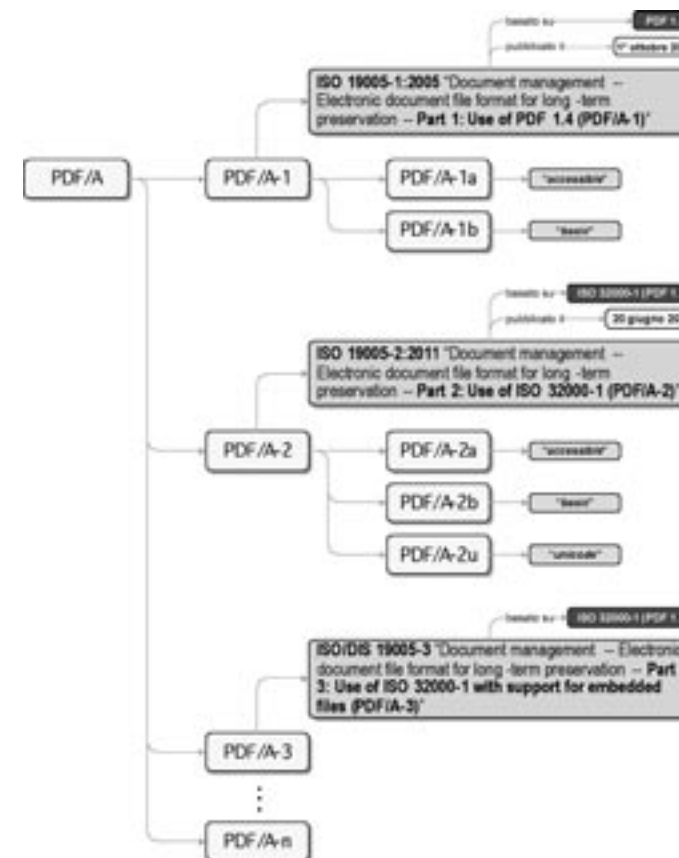


Fig. 1. Le parti che compongono lo standard ISO 19005.

contiene le specifiche del formato PDF versione 1.4, utilizzato in Adobe Acrobat 5. Analogamente il *PDF Reference 1.5* contiene le specifiche del formato PDF versione 1.5, utilizzato in Adobe Acrobat 6, e così via. I *PDF Reference* sono disponibili gratuitamente sul sito di Adobe Systems. Ad esempio, l'ultima versione, la 1.7, è disponibile all'indirizzo <http://www.adobe.com/devnet/pdf/pdf_reference.html>.

⁴ Il processo di pubblicazione è stato piuttosto veloce: il 30 novembre 2010 il Comitato tecnico 171, responsabile dello sviluppo dello standard ISO 19005, si era riunito in Ottawa e aveva ratificato la seconda parte dello standard (ISO 19005-2); era seguito il processo di assicurazione della qualità ed il processo di pubblicazione presso il Segretariato centrale dell'ISO a Ginevra fino a giungere, il 20 giugno 2011, al traguardo finale, la pubblicazione dello standard.

⁵ Le norme citate nel presente articolo sono disponibili, a pagamento, sul sito dell'ISO <<http://www.iso.org>>.

⁶ Lo sviluppo è condotto dal Sottocomitato 2 Application issues del Comitato Tecnico 171 Document management applications (ISO/TC171/SC2 – *Document management applications – Application issues*) in seno all'ISO.

1.2 Perché una “seconda parte”?

Il PDF/A-1 si basa sul *PDF Reference 1.4*, pubblicato da Adobe Systems nel 2001 e rappresenta un sottoinsieme (*subset*)⁷ del formato PDF 1.4 (si veda la Figura 2), una sua versione “limitata” che è stata privata di tutte quelle caratteristiche che non sono adatte per la conservazione a lungo termine dei documenti informatici⁸. Ma dalla pubblicazione della norma ISO 19005-1, che risale al 2005, molte

⁷ In letteratura si utilizza anche il termine “profilo”: il PDF/A costituisce un particolare *profilo* del PDF. Il termine è utilizzato, ad esempio, anche per il PDF/X, un *profilo* del PDF utilizzato nel settore della stampa e delle arti grafiche.

⁸ Secondo la definizione fornita dalla Library of Congress il PDF/A è “a constrained form of Adobe PDF version 1.4 intended to be suitable for long-term preservation of page-oriented documents for which PDF is already being used in practice”. Si veda <<http://digitalpreservation.gov/formats/fdd/fdd000125.shtml>>.

cose sono cambiate. Il formato PDF ha continuato a svilupparsi introducendo nuove interessanti funzionalità, molte delle quali non sono risultate in contrasto con lo scopo del formato PDF/A di garantire la conservazione nel tempo dei documenti elettronici prodotti secondo tale formato. Nel frattempo, nel 2008, le specifiche del PDF 1.7 erano state riconosciute standard ISO con la pubblicazione della norma ISO 32000-1:2008 *Document management – Portable document format – Part 1: PDF 1.7*; di conseguenza il formato PDF, nella versione 1.7, non è più una specifica controllata da Adobe Systems ma è diventato uno standard riconosciuto a livello internazionale. Contemporaneamente, da più parti si era cominciato ad avvertire l'esigenza di rilasciare una nuova parte della norma ISO 19005 basata non più su una specifica pubblicata da Adobe Systems ma su uno standard internazionale (quindi: uno standard ISO basato su uno standard ISO). I tempi erano perciò maturi per una evoluzione del formato PDF/A ed il Comitato tecnico ISO/TC171/SC2 ha cominciato a lavorare sulla introduzione di alcune nuove funzionalità nel formato, prendendo questa volta come base il PDF 1.7 (come definito dalla norma ISO 32000-1:2008). Questo ha portato all'inserimento, nel nuovo formato denominato PDF/A-2, di alcune utili funzionalità che erano già state gradualmente introdotte nelle versioni 1.5, 1.6 e 1.7 del formato PDF.

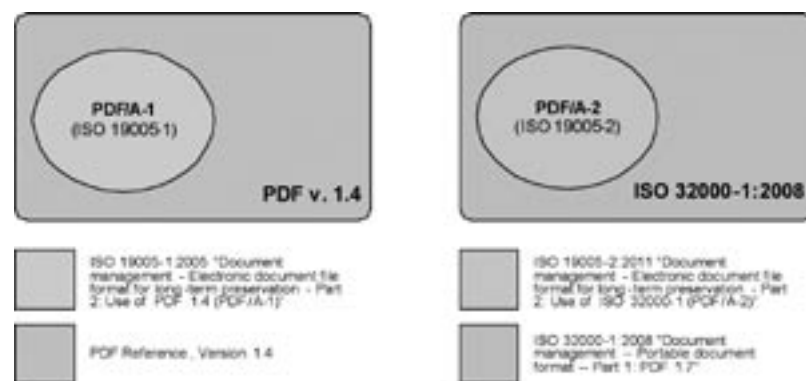


Fig. 2. Il PDF/A-2 è un subset dell'ISO 32000-1, così come il PDF/A-1 è un *subset* del PDF 1.4.

1.3 Le novità del PDF/A-2

Come il PDF/A-1, anche il PDF/A-2 si prefigge lo scopo di garantire la conservazione a lungo termine dei documenti elettronici codificati secondo tale formato; quindi esso cerca di assicurare i seguenti requisiti:

- l'indipendenza dal dispositivo (*device independence*): le stesse informazioni e lo stesso aspetto visivo del documento devono essere garantite per tutti i dispositivi (entro i limiti imposti dal dispositivo);

- l'auto-contenimento (*self-containment*): tutte le informazioni necessarie per rappresentare correttamente il documento devono essere contenute nel *file* stesso;
- l'auto-documentazione (*self-documentation*): il *file* deve contenere al suo interno i metadati necessari per la descrizione del documento.

Il criterio che ha guidato nella definizione del PDF/A-2 è stato quello di accogliere tutte le caratteristiche del formato PDF 1.7 che non interferiscono con gli obiettivi di conservazione a lungo termine dei documenti elettronici e di vietare, invece, tutte quelle caratteristiche che vanno in contrasto con questi obiettivi, in maniera da preservare nel tempo la rappresentazione visiva dei documenti.

Tuttavia, rispetto al PDF/A-1, alcune limitazioni rimangono in essere mentre altre vengono superate. Ad esempio, esattamente come avviene per i file PDF/A-1, un file PDF/A-2 deve incorporare tutti i *font* utilizzati nel documento (ed essi devono essere, inoltre, legalmente incorporabili) per garantire una rappresentazione del testo sempre uguale; gli spazi dei colori devono essere specificati in una maniera indipendente dal dispositivo per assicurare la possibilità di rappresentare in modi sempre uguali i colori utilizzati nel documento; non devono essere presenti contenuti "recuperabili" attraverso collegamenti ipertestuali (*link*) verso fonti esterne⁹; non possono essere incorporati contenuti audio e video; l'uso del codice JavaScript è assolutamente vietato ed è consentito solo un limitato insieme di *action* (quelli per i quali non esiste alcun rischio che possa essere modificato il contenuto dei documenti); non può essere utilizzata la crittografia per cifrare il contenuto e devono essere obbligatoriamente utilizzati metadati standard.

A differenza, invece, del PDF/A-1, il PDF/A-2 raggiunge una maggiore efficienza nella memorizzazione delle immagini grazie all'adozione dell'algoritmo di compressione JPEG2000, supporta gli effetti di trasparenza e l'impiego dei *layer*, permette l'incorporamento di *font* OpenType e la possibilità di sottoscrivere i documenti mediante firme elettroniche conformi allo standard PAdES (*PDF Advanced Electronic Signatures*). In aggiunta, il PDF/A-2 offre la possibilità di incorporare file PDF/A (sia PDF/A-1 che PDF/A-2) all'interno di file PDF/A-2, permettendo l'archiviazione in un unico file di insiemi di documenti, ciascuno dei quali mantiene tuttavia la propria autonomia. Il *focus* sui contenuti statici continua, tuttavia, ad escludere la possibilità di inserire contenuti multimediali all'interno di un *file* in formato PDF/A.

⁹ Questo non implica, tuttavia, che in un *file* PDF/A-1 o PDF/A-2 non possano essere presenti meccanismi di interazione; al contrario, possono essere presenti *link* di navigazione, sia che essi puntino verso altre parti all'interno dello stesso *file*, sia che essi puntino verso *file* esterni o altre destinazioni. Quello che, invece, rimane vietato è l'utilizzo dei *link* per il "recupero" di parti del documento (ad esempio, una immagine) da sorgenti esterne al documento e la successiva visualizzazione all'interno del documento stesso.

I paragrafi che seguono illustrano più nel dettaglio le principali nuove funzionalità del PDF/A-2 che potrebbero avere un notevole impatto sull'utilizzo di questo formato per l'archiviazione e la conservazione dei documenti elettronici¹⁰.

1.3.1 Supporto alle collezioni di file PDF/A

La prima interessante novità introdotta dal PDF/A-2 è costituita dal fatto che permette di incorporare altri file PDF/A (sia PDF/A-1 che PDF/A-2) all'interno di un *file* PDF/A-2, in maniera da realizzare *collezioni* di file PDF/A¹¹. In questo modo un singolo *file* PDF/A-2 può essere utilizzato come una sorta di “contenitore” di altri file PDF/A e ciò risulta utile quando diversi *file* PDF/A devono essere conservati insieme pur rimanendo *file* autonomi (cioè senza essere uniti in un unico *file* PDF/A). Ad esempio, è possibile convertire una e-mail con i suoi allegati in un *file* PDF/A-2 nel quale gli allegati sono incorporati anch'essi come file PDF/A. Oppure è possibile raccogliere in un unico *file* PDF/A-2 che funge da “contenitore” virtuale diversi file PDF/A che sono stati individualmente sottoscritti con firma elettronica, evitando così che la loro fusione in un unico file comprometta la validità delle firme. Ancora, più semplicemente, a partire da una cartella contenente più documenti PDF/A è possibile creare un unico file PDF/A-2 con incorporati tanti file PDF/A quanti sono i documenti di partenza. Il sistema prevede anche la possibilità di aggiungere o rimuovere elementi (senza alterare la conformità del file “contenitore” alle specifiche del PDF/A-2) ed un meccanismo di navigazione tra i vari elementi.

1.3.2 Supporto all'utilizzo degli oggetti trasparenti

Il PDF/A-2 prevede la possibilità di utilizzare gli oggetti trasparenti. In realtà, questa possibilità era già presente nella specifica del PDF 1.4, ma non era stata considerata una funzionalità ammissibile nel PDF/A-1 soprattutto perché si era ritenuto che la tecnologia relativa non fosse ancora sufficientemente matura. Inoltre, in tale specifica alcuni dettagli tecnici non erano ancora del tutto chiari e ciò avrebbe potuto portare, in determinate circostanze, ad una rappresentazione degli oggetti trasparenti non sempre costante.

¹⁰ Si è fatto qui riferimento all'articolo di Olaf Drümmer, *What is PDF/A-2?* disponibile all'indirizzo <http://www.druemmer.com/_wp/?p=27> e all'articolo di Duff Johnson, *PDF/A-2 Ratified*, in “PDF/A Competence Center Newsletter”, 2010, 19, disponibile all'indirizzo <<http://www.pdfa.org/2010/12/pdfa-competence-center-newsletter-issue-19/>>.

¹¹ Si tratta di un concetto simile a quello del “portfolio” previsto nel formato PDF, il quale, tuttavia, permette di incorporare non solo *file* PDF/A ma anche altre tipologie di *file* (ad esempio, documenti di Microsoft Word, fogli di lavoro di Microsoft Excel, immagini, etc.); evidentemente questa possibilità mal si sposa con gli obiettivi di conservazione a lungo termine.



Fig 3. Esempio di documento contenente trasparenze (si vedano le parti evidenziate).

Oggi, invece, la tecnologia che sta alla base degli oggetti trasparenti è decisamente maturata e lo standard PDF 1.7 ne definisce ora con sufficiente precisione i dettagli tecnici. Di conseguenza gli effetti di trasparenza possono essere implementati in maniera sempre coerente e non vi sono più ostacoli al suo utilizzo all'interno di un documento in formato il PDF/A-2¹².

La trasparenza viene utilizzata in diversi modi nei file PDF: un tipico esempio è costituito dalle annotazioni sui documenti di testo effettuate mediante lo strumento “evidenziatore” (nei vari colori giallo, verde, etc.; si veda la Figura 3); anche le presentazioni utilizzate nei convegni spesso fanno uso di effetti codificati per mezzo della trasparenza, così come alcuni effetti utilizzati nel settore del disegno grafico.

¹² In precedenza, con il formato PDF/A-1, che esclude espressamente la possibilità di utilizzare la trasparenza, tutti gli effetti di trasparenza contenuti in un documento dovevano essere “appiattiti” prima che il documento stesso potesse essere convertito in un file PDF/A-1.

1.3.3 Supporto al JPEG2000

Il PDF/A-2, così come il formato PDF 1.7, consente l'utilizzo del JPEG2000 come formato per la compressione delle immagini incorporate in un file PDF/A-2. Il JPEG2000 è anch'esso uno standard ISO (ISO/IEC 15444¹³) creato nell'anno 2000 dal medesimo gruppo (il *Joint Photographic Experts Group*) che ha creato l'algoritmo di compressione per immagini JPEG con l'obiettivo a lungo termine di prenderne il posto. Adotta un algoritmo molto efficiente che permette rapporti di compressione molto elevati senza la produzione degli artefatti tipici del formato JPEG. In certi ambiti di applicazione è molto più efficiente dell'algoritmo utilizzato dal formato JPEG, consentendo interessanti riduzioni nelle dimensioni dei file che si ottengono; un tipico caso in cui la compressione JPEG2000 risulta molto utile è quello delle immagini ottenute mediante scansione.

La maggiore capacità di compressione dell'algoritmo JPEG2000 rispetto al JPEG è una caratteristica che si presenta molto interessante per tutto l'ambito della dematerializzazione dei documenti analogici. Infatti, una pratica oggi piuttosto comune (soprattutto presso talune pubbliche amministrazioni ed aziende) consiste nel produrre file PDF/A a partire da scansioni di documenti di testo sottoscritti con firma autografa. Questi file PDF vengono ampiamente utilizzati, ad esempio per la pubblicazione all'albo pretorio on-line o per lo scambio di note, circolari ed altre tipologie di documenti attraverso la posta elettronica. Purtroppo, questo procedimento di creazione di file PDF/A mediante scansione di documenti analogici finisce con il produrre file di dimensioni piuttosto elevate (nel gergo comune si parla di file "pesanti"). Ora, grazie all'adozione da parte del PDF/A-2 dell'algoritmo di compressione JPEG2000, più efficiente rispetto all'algoritmo JPEG precedente adottato dal PDF/A-1, è possibile produrre file PDF/A che hanno una dimensione minore (e, quindi, un "peso" minore). Quindi, in casi come questo, l'adozione del PDF/A-2 rappresenta senz'altro la scelta migliore.

¹³ Il formato per immagini JPEG2000 è definito dalla norma ISO/IEC 15444, la quale è strutturata attualmente in 12 parti numerate da 1 a 13 (la parte n. 7 è stata ritirata): ISO/IEC 15444-1:2004 *Information technology – JPEG 2000 image coding system: Core coding system*; ISO/IEC 15444-2:2004 *Information technology – JPEG 2000 image coding system: Extensions*; ISO/IEC 15444-3:2007 *Information technology – JPEG 2000 image coding system: Motion JPEG 2000*; ISO/IEC 15444-4:2004 *Information technology – JPEG 2000 image coding system: Conformance testing*; ISO/IEC 15444-5:2003 *Information technology – JPEG 2000 image coding system: Reference software*; ISO/IEC 15444-6:2003 *Information technology – JPEG 2000 image coding system – Part 6: Compound image file format*; ISO/IEC 15444-8:2007 *Information technology – JPEG 2000 image coding system: Secure JPEG 2000*; ISO/IEC 15444-8:2007/Amd 1:2008 *File format security*; ISO/IEC 15444-9:2005 *Information technology – JPEG 2000 image coding system: Interactivity tools, APIs and protocols*; ISO/IEC 15444-10:2008 *Information technology – JPEG 2000 image coding system: Extensions for three-dimensional data*; ISO/IEC 15444-11:2007 *Information technology – JPEG 2000 image coding system: Wireless*; ISO/IEC 15444-12:2008 *Information technology – JPEG 2000 image coding system – Part 12: ISO base media file format*; ISO/IEC 15444-13:2008 *Information technology – JPEG 2000 image coding system: An entry level JPEG 2000 encoder*.

1.3.4 Supporto alla gestione dei layer

Il PDF/A-2 introduce la gestione dei *layer* (livelli)¹⁴, grazie ai quali è possibile scegliere quali parti del file visualizzare e quali nascondere. Nella pratica i *layer* sono molto utilizzati nelle comunità professionali degli ingegneri e degli architetti, che ne fanno uso nei complessi disegni tecnici nei quali può essere utile mostrare o nascondere alcuni aspetti del disegno per concentrarsi più facilmente sulle informazioni rimanenti. Ad esempio, nel caso di un elaborato grafico relativo al progetto di una casa, potrebbe essere utile disattivare la visualizzazione del *layer* relativo all'impianto elettrico quando ci si vuole concentrare sull'impianto idrico, e viceversa. I *layer* sono utilizzati anche per le pubblicazioni multi-lingue e consentono la visualizzazione del testo in una sola lingua alla volta. In questo modo è possibile realizzare un unico documento contenente un *layer* per ciascuna lingua (anziché più documenti ciascuno in una determinata lingua), con una maggiore usabilità ed una evidente ottimizzazione della dimensione dei file dal momento che non occorre produrre tanti documenti quante sono le lingue utilizzate e inoltre gli eventuali elementi grafici (ad esempio, le immagini) presenti nel documento non necessitano di essere duplicati in più documenti, uno per ciascuna lingua.



Fig. 4. Esempio di documento PDF/A-2 multi-lingue.

¹⁴ Da un punto di vista tecnico nelle specifiche del formato PDF i livelli sono denominati "Optional Content Group (OCG)": ad esempio, nella specifica PDF Reference 1.5 si legge "Optional content refers to sections of content in a PDF document that can be selectively viewed or hidden by document authors or consumers. This capability is useful in items such as CAD drawings, layered artwork, maps, and multi-language documents". Tuttavia, la maggior parte delle applicazioni software per la gestione di file PDF utilizza il più comune termine "layer".

Si pensi, ad esempio, alle pubblicazioni ufficiali dell'Unione europea che, anziché essere rese disponibili in più lingue (solitamente un *file* PDF per ogni lingua), possono essere rese disponibili sotto forma di un unico file PDF/A-2 che fa uso di un *layer* per ogni lingua. In Figura 4 è riportato un esempio di una possibile applicazione dell'utilizzo dei *layer* alla pubblicazione in formato elettronico della Gazzetta ufficiale dell'Unione europea: è sufficiente utilizzare un solo documento PDF/A-2 contenente tanti *layer* quante sono le lingue nelle quali deve essere pubblicata la Gazzetta ufficiale (nell'esempio, per semplicità, sono stati utilizzati solamente tre *layer*, corrispondenti alle lingue inglese, italiano e francese), anziché tanti file PDF, uno per ciascuna lingua.

Questa nuova funzionalità aggiunta al PDF/A-2, se da un lato può essere vista positivamente in quanto la gestione e la conservazione di un unico file in luogo di due o più file è sicuramente più semplice e consente, quindi, di raggiungere livelli maggiori di efficienza, dall'altro deve essere valutata attentamente per quanto riguarda il requisito della "staticità" del documento in merito al quale, a ben vedere, l'utilizzo dei *layer* potrebbe creare qualche problema, andando ad influire sull'univocità della rappresentazione del documento stesso. Si tratta, pertanto, di un aspetto che merita di essere ulteriormente approfondito.

1.3.5 Incorporazione dei font OpenType

Il formato PDF/A-2 consente l'utilizzo e l'incorporazione nei documenti dei *font* OpenType¹⁵, caratteristica questa introdotta dalla versione 1.7 del formato PDF. Mentre con il PDF/A-1 i *font* OpenType utilizzati per produrre documenti di testo non potevano essere incorporati direttamente e dovevano necessariamente essere convertiti prima in *font* PostScript Type 1 o TrueType, con il PDF/A-2 i *font* OpenType possono essere incorporati direttamente senza alcuna conversione. Si tratta, anche in questo caso, di un miglioramento notevole verso l'obiettivo della standardizzazione completa (non solo del PDF/A ma anche delle tecnologie su cui si basa), dal momento che dall'anno 2009 anche i *font* OpenType sono stati riconosciuti

¹⁵ L'OpenType è un formato di carattere tipografico (*font*) di tipo vettoriale sviluppato congiuntamente da Microsoft Corporation e Adobe Systems alla fine degli anni '90 e concepito quale successore dei preesistenti formati di carattere tipografico che, sino ad allora, si contendevano i favori del mercato: TrueType (sviluppato da Apple e da Microsoft Corporation) e PostScript Type 1 (creato da Adobe Systems). Il primo rilascio di un numero significativo di *font* si ebbe nel 2000-2001. Adobe Systems ha completato la conversione della sua intera libreria di *font* in OpenType all'incirca alla fine del 2002. I *font* OpenType presentano molte caratteristiche rilevanti: sono indipendenti dalla piattaforma (*cross-platform*), perciò è possibile utilizzare lo stesso *font* sia su sistemi Macintosh che Windows; possono includere "caratteri intelligenti", come frazioni automatiche, caratteri speciali e selezioni intelligenti della forma del carattere (il cosiddetto *glifo*) in base alla loro posizione. Ad esempio, con un *font* calligrafico la lettera "t" ha una determinata forma quando compare all'inizio di una parola, un'altra forma quando compare nel mezzo di una parola (in maniera che si "leggi" agli altri caratteri) e può anche presentare un'ulteriore forma (il cosiddetto "svolazzo") se posto in prossimità della fine di una parola.

standard sia dall'International Organization for Standardization (ISO) che dall'International Electrotechnical Commission (IEC) con la norma ISO/IEC 14496-22:2009 *Information technology – Coding of audio-visual objects – Part 22: Open Font Format*.

1.3.6 Supporto migliorato alle annotazioni

Tra la versione 1.4 e quella 1.7 del PDF è stato introdotto nel formato PDF un certo numero di nuove annotazioni, utilizzabili per l'aggiunta di commenti (come i riquadri, gli ovali, le nuvolette, i poligoni, le linee poligonali, etc.) o di marcature (come le filigrane, i timbri, le note di revisione, etc.); alcuni di questi rimangono tutt'ora proibiti, mentre altri possono ora essere inseriti nei *file* PDF/A-2.

1.3.7 Supporto per le firme elettroniche secondo lo standard PaDES

La sottoscrizione di un documento mediante firma elettronica era già prevista dal PDF/A-1. Il PDF/A-2 recepisce le nuove disposizioni previste dallo standard PAdES (*PDF Advanced Electronic Signatures*)¹⁶, che specifica appositi profili per l'utilizzo, ai sensi della direttiva dell'Unione europea 1999/93/CE, della firma elettronica avanzata all'interno dei documenti PDF.

Il formato di firma PAdES costituisce una estensione dello standard di firma PDF già definito da Adobe ed anche se non è a tutt'oggi il metodo di firma più utilizzato, la larga diffusione del formato PDF rende il PAdES un formato da non trascurare per diffondere l'utilizzo della firma elettronica su bacini di utenza molto vasti. Inoltre, a differenza di altri formati di firma, nel caso dell'utilizzo della firma PAdES il documento da sottoscrivere non subisce operazioni di trasformazione di formato né necessita di essere sottoposto ad operazioni di imbustamento/sbustamento; inoltre, essendo contenuta all'interno del PDF, può essere visualizzato da un qualunque lettore di file PDF (ad esempio, il gratuito Acrobat Reader), senza la necessità di ricorrere all'applicazione che ha generato il documento (ad esempio, Microsoft Word o Excel). Il supporto alla firma elettronica secondo lo standard PAdES rappresenta, quindi, un ulteriore passo sulla strada di una maggiore adozione del PDF/A come formato per la conservazione di documenti digitali sottoscritti con firma elettronica.

¹⁶ Lo standard PAdES è definito dalla norma ETSI TS 102 778, la quale si compone di cinque parti: Part 1: "PAdES Overview – a framework document for PAdES"; Part 2: "PAdES Basic – Profile based on ISO 32000-1"; Part 3: "PAdES Enhanced – PAdES-Basic Electronic Signatures and PAdES-Explicit Policy Electronic Signatures Profiles"; Part 4: "PAdES Long Term – PAdES-Long Term Validation Profile"; Part 5: "PAdES for XML Content – Profiles for XAdES signatures of XML content in PDF files". Tali standard sono disponibili dalla pagina "ETSI Publications Download Area" all'indirizzo <<http://pda.etsi.org/pda/queryform.asp>>.

1.3.8 Maggiori dimensioni dei documenti

Le dimensioni di un documento in formato PDF/A-1 sono al massimo di 5,08 x 5,08 metri; si tratta di dimensioni limitate che impongono tutta una serie di restrizioni difficilmente superabili: si pensi alle stampe di grande formato, alla documentazione cartografica (ad esempio, le mappe prodotte dai sistemi GIS, Geographical Information System), etc. Con il PDF/A-2 le dimensioni massime vengono portate a 381 x 381 km, dimensioni che sono sicuramente più che sufficienti per qualsiasi tipologia di lavoro. Anche in questo caso questa nuova caratteristica introdotta dal PDF/A-2 deve essere valutata molto positivamente perché consente di estendere i benefici del formato PDF/A a tutti quei settori (come quelli della progettazione ingegneristica ed architettonica, quelli della grafica e della stampa professionale, etc.) che fino ad oggi erano rimasti esclusi dal suo campo di applicazione per via delle limitate dimensioni massime previste dal PDF/A-1.

Per ulteriori approfondimenti sul formato PDF/A ed in particolare sul formato PDF/A-2 si rimanda al sito web del PDF/A Competence Center <<http://www.pdfa.org>>, l'associazione internazionale che collabora allo sviluppo dello standard e si occupa della sua promozione.

1.4 I livelli di conformità del PDF/A-2

Il PDF/A-1 prevede solo due possibili livelli di conformità allo standard (i livelli PDF/A-1a e PDF/A-1b); il PDF/A-2 aggiunge un terzo livello di conformità, il PDF/A-2u (unicode), portando a tre il numero dei possibili livelli di conformità (i primi due dei quali sono analoghi a quelli previsti per il formato PDF/A-1). Più precisamente:

- il PDF/A-2b (*basic*) è il livello di conformità minima; un file PDF/A-2b è più semplice da generare ma assicura solamente la conservazione visiva dell'aspetto del documento¹⁷. In altre parole, questa versione assicura che il documento venga visualizzato correttamente ma non garantisce, ad esempio, che il testo estratto sia estraibile o che sia possibile interpretare semanticamente il suo contenuto.
- il PDF/A-2a (*accessible*) è il livello di conformità completa, che assicura la conformità a tutti i requisiti dell'ISO 19005-2; un file PDF/A-2 contiene sufficienti informazioni (sotto forma sia di testo di supporto in formato Unicode che di tag) da assicurare la possibilità di accedere in maniera avanzata ai suoi contenuti, di recuperarli in maniera strutturata (specialmente con riferimento alle parti di testo) e di disporre di funzioni di ricerca¹⁸. Questo è un presupposto importante, ad esempio, ai fini del riutilizzo del testo su altre piattaforme o per l'esecuzione

¹⁷ Ad esempio, un documento su carta, acquisito, mediante il processo di scansione, in formato immagine e successivamente convertito in formato PDF, rispetta i requisiti del livello di conformità PDF/A-2b, ma il testo non risulta ovviamente estraibile (se non sottoponendolo ad un processo di OCR).

¹⁸ Grazie ai *tag* è possibile fornire informazioni sulla struttura del file e sulla semantica del testo, in maniera da rendere possibile la conservazione della sua struttura logica e dell'ordine di lettura.

delle operazioni di migrazione che si renderanno necessarie nel futuro. La struttura di un file PDF/A-2a, infatti, è tale che il riutilizzo di testo – ad esempio, mediante una funzione di esportazione – non presenta generalmente difficoltà (evitando problemi quali paragrafi disallineati, parole troncate, etc.).

- il PDF/A-2u (unicode) è il livello di conformità intermedio e richiede il soddisfacimento di requisiti meno completi di quelli di livello PDF/A-2a ma più completi rispetto al PDF/A-2b: la “u” sta a significare che il testo presente nel documento è codificato in Unicode¹⁹. Inoltre, questo livello richiede che i *file* così ottenuti siano ricercabili e che il contenuto del testo possa essere estratto e copiato. Il livello di conformità “u” è una versione semplificata del livello di conformità “a”, poiché offre i vantaggi di Unicode relativamente alla possibilità di effettuare ricerche sul testo e di copiare i contenuti senza tuttavia richiedere il soddisfacimento dei requisiti di struttura logica previsti per il livello di conformità “a”. Il livello di conformità “a” è, ovviamente, preferibile ma richiede uno sforzo maggiore durante il processo di generazione e non sempre è raggiungibile. Ad esempio, i documenti acquisiti mediante scansione ottica mancano di qualsiasi informazione sulla struttura dei contenuti e, quindi, non possono raggiungere il livello di conformità “a”. Allo stesso modo i file PDF creati da applicazioni che non supportano la creazione di file PDF con tag (i cosiddetti “tagged PDF”) non possono essere prodotti al livello di conformità “a”. Tuttavia, tali file PDF possono ancora essere prodotti in maniera da garantire che tutti i testi siano codificati in Unicode. Si pensi, ad esempio, ai documenti in formato PDF acquisiti mediante scansione ottica e resi “ricercabili” mediante un software di riconoscimento ottico dei caratteri (OCR, *Optical Character Recognition*).

Trattandosi di condizioni più stringenti, un file conforme alle specifiche del livello “a” è conforme anche a quelle dei livelli “b” ed “u”. Non vale ovviamente il viceversa: un file conforme alle specifiche del livello “b” non necessariamente è conforme alle specifiche del livello “u” o a quelle del livello “a” (si veda la Figura 5, che rappresenta graficamente le relazioni tra il PDF/A2-a, il PDF/A2-b e il PDF/A-2u).

¹⁹ Unicode è un sistema di codifica dei caratteri utilizzato per la scrittura di testi in maniera indipendente dalla lingua, dalla piattaforma informatica e dal programma utilizzati. È descritto nello standard ISO/IEC 10646: *Information technology – Universal Multiple-Octet Coded Character Set (UCS) (UNICODE)*. Il supporto ad Unicode è stato ritenuto così importante da introdurre nel PDF/A-2 l'ulteriore livello di conformità “u”.



Fig. 5. La relazione tra i due livelli di conformità del PDF/A-1 (a sinistra) e tra i tre livelli di conformità del PDF/A-2 (a destra). Si noti che il PDF/A-2 introduce un livello di conformità intermedio (il PDF/A-2u).

1.5 La compatibilità all'indietro

Il PDF/A-2 è compatibile all'indietro con la precedente versione PDF/A-1; questo significa che qualunque documento conforme alle specifiche del PDF/A-1 (ovvero, "PDF/A-1 compliant") è automaticamente conforme anche alle specifiche del PDF/A-2 (ovvero è anche "PDF/A-2 compliant"). Questo significa che non è assolutamente necessario prevedere alcuna operazione di conversione dei file già esistenti in formato PDF/A-1 nel formato PDF/A-2²⁰; infatti, se si sottopone a conversione nel formato PDF/A-2 un file che è PDF/A-1 compliant, il file non subisce alcuna modifica, perché qualunque caratteristica che non rispetta le specifiche PDF/A-2 è già stata rimossa nella precedente conversione che lo ha trasformato in file PDF/A-1.

Il contrario non è vero: un documento conforme alle specifiche del formato PDF/A-2 non necessariamente è conforme anche alle specifiche del formato PDF/A-1²¹. Ad esempio, un documento PDF contenente immagini compresse secondo l'algoritmo JPEG2000 può essere conforme alle specifiche del PDF/A-2 ma non lo sarà sicuramente rispetto alle specifiche del PDF/A-1 (che vieta espressamente l'utilizzo della compressione JPEG2000).

La relazione tra le parti del PDF/A già pubblicate come standard ISO e quelle che seguiranno in futuro è schematizzata nella Figura 6, dalla quale si osserva che un file conforme alle specifiche del formato PDF/A-1 è sicuramente conforme anche alle specifiche del formato PDF/A-2, del PDF/A-3, e così via.

²⁰ È auspicabile che anche nel futuro venga sempre garantita la compatibilità con le versioni successive del formato (*forward compatibility*): in questo modo i file che sono conformi alle specifiche PDF/A-2 risulteranno conformi anche alle successive specifiche PDF/A-3, PDF/A-4, e così via. Ovviamente non sarà possibile garantire il contrario.

²¹ Quindi per il formato PDF/A è assicurata la compatibilità in avanti (*forward compatibility*) ma non quella all'indietro (*backward compatibility*).

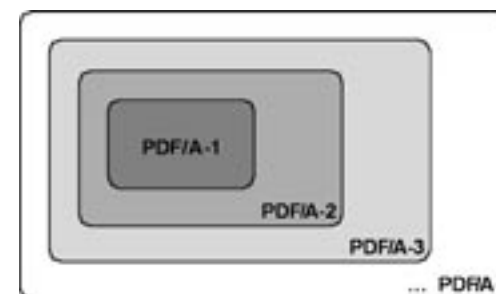


Fig. 6. La relazione tra le varie parti del PDF/A.

1.6 L'identificazione dei file

I nomi dei file conformi allo standard PDF/A-2 (così come quelli dei file conformi allo standard PDF/A-1) hanno come estensione .pdf (esattamente la stessa dei file PDF) dal momento che gli standard ISO 19005-1 e ISO 19005-2 non richiedono che debba essere utilizzata una differente estensione per distinguere tali formati dal formato PDF²². Tuttavia in taluni ambienti è ormai invalsa una consuetudine che si ritiene di particolare interesse perché consente di distinguere, visivamente, i file PDF/A da quelli che sono "semplici" PDF. Tale consuetudine consiste nell'aggiungere al nome del file un suffisso che richiama il livello di conformità al quale il file risulta corrispondente.

In particolare:

- il suffisso "_1a" indica che il file è conforme al livello di conformità PDF/A-1a;
- il suffisso "_1b" indica che il file è conforme al livello di conformità PDF/A-1b;
- il suffisso "_2a" indica che il file è conforme al livello di conformità PDF/A-2a;
- il suffisso "_2u" indica che il file è conforme al livello di conformità PDF/A-2u;
- il suffisso "_2b" indica che il file è conforme al livello di conformità PDF/A-2b;

Ad esempio, con la denominazione "Circolare_1a.pdf" si vuole indicare che si tratta di un file PDF corrispondente al livello di conformità PDF/A-1a, mentre con la denominazione "Circolare_2b.pdf" si fa riferimento ad un file PDF corrispondente al livello di conformità PDF/A-2b.

²² Sarà compito del programma di visualizzazione dei file PDF/A individuare il formato corretto. Si noti che, anche per altri formati, non è possibile distinguere le varie versioni tramite l'estensione del nome del file. Per fare un esempio, le diverse versioni del software di videoscrittura Microsoft Word fino alla versione 2003 producono file che hanno tutti la ben nota estensione .doc ma, di fatto, fanno riferimento alle differenti versioni dello stesso formato che si sono avvicendate nel tempo. È compito del software comprendere di quale versione si tratti. Allo stesso modo, le varie versioni del formato PDF (dalla 1.0 alla 1.7 attuale), compresi i diversi profili PDF/A-1 e PDF/A-2, producono file che sono identificati tutti dalla stessa estensione .pdf.

È bene osservare che il fatto che il nome del file contenga, ad esempio, il suffisso “_2b” non garantisce, ovviamente, sul fatto che il file sia effettivamente conforme alle specifiche del formato PDF/A-2b²³.

Per avere la certezza che si tratti effettivamente di un file PDF/A (e, specificatamente, di un file conforme ad un certo profilo) occorre ricorrere ad apposite applicazioni software, i cosiddetti *validatori*²⁴ PDF/A, che consentono di verificare la conformità di un file PDF ai vari profili ad oggi esistenti (PDF/A-1a, PDF/A-1b, PDF/A-2a, PDF/A-2u e PDF/A-2b).

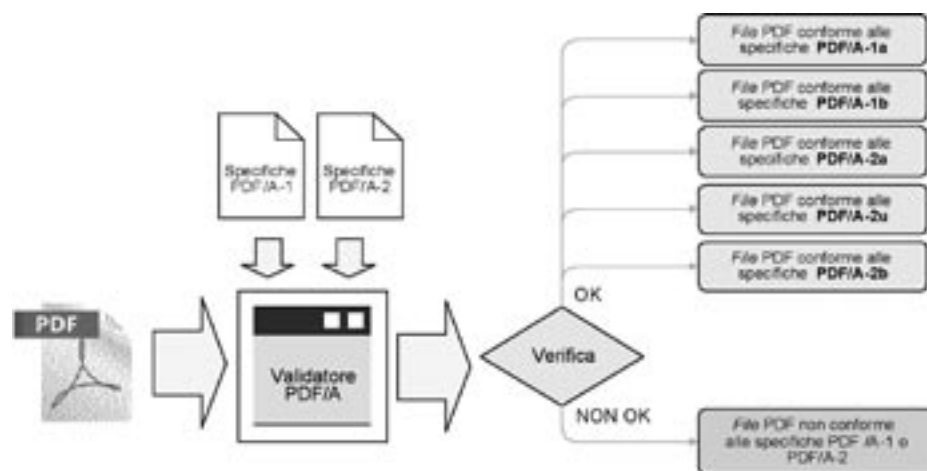


Fig. 7. Esempio di validatore PDF/A.

²³ Infatti, potrebbe trattarsi di un semplice file PDF e non PDF/A, oppure di un file che non è neppure in formato PDF ma al quale è stato associato un nome ed una estensione tali da ricondurlo (erroneamente) ad un file in formato PDF/A. Ricordiamo, infatti, che il nome di un file e la sua estensione vengono attribuiti dal soggetto che li produce, il quale ha piena facoltà di attribuirgli qualsiasi denominazione (anche non corrispondente al formato del file). Per un approfondimento si rimanda a S. Pigliapoco, S. Allegrezza, *Produzione e conservazione del documento digitale. Requisiti e standard per i formati elettronici*. Volume I, Edizioni EUM, Macerata, 2008. Si veda anche M. Guercio, *Archivistica informatica. I documenti in ambiente digitale*, Nuova edizione, Carocci editore, Roma, 2010 (in particolare il paragrafo *Il ruolo dei formati*, pagg. 180 -189).

²⁴ In generale, un *validatore* consente di verificare se un *file* in un determinato formato è conforme ad uno specifico profilo di quel formato.

È interessante analizzarne il meccanismo di funzionamento secondo lo schema rappresentato nella Figura 7. Il validatore riceve in input un file PDF qualsiasi ed effettua una verifica della sua conformità alle specifiche dei formati PDF/A-1 o PDF/A-2. L'esito di tale verifica può produrre i seguenti possibili risultati:

- conformità alle specifiche del PDF/A-1a (e, di conseguenza, anche a quelle del PDF/A-1b, nonché a quelle del PDF/A-2);
- conformità alle specifiche del PDF/A-1b (e, di conseguenza, anche a quelle del PDF/A-2);
- conformità alle specifiche del PDF/A-2a (e, di conseguenza, anche a quelle del PDF/A-2u e PDF/A-2b);
- conformità alle specifiche del PDF/A-2u (e, di conseguenza, anche a quelle del PDF/A-2b);
- conformità alle specifiche del PDF/A-2b;
- non conformità né alle specifiche del PDF/A-1 né a quelle del PDF/A-2.

All'interno della versione più recente di Adobe Acrobat (la versione Acrobat X nel momento in cui si scrive) è incluso un validatore sia nei confronti del formato PDF/A-1 che nei confronti del formato PDF/A-2, ma esistono anche validatori di terze parti che svolgono in maniera analoga queste funzioni.

1.7 Conclusioni

Il nuovo formato PDF/A-2 introduce interessanti novità che dovrebbero essere valutate attentamente da tutti coloro che si occupano di archiviazione e conservazione digitale (archivisti informatici, record manager, responsabili della conservazione, responsabili del Servizio per la tenuta del protocollo informatico, la gestione dei flussi documentali e degli archivi, etc.).

Rispetto al suo predecessore PDF/A-1, il formato PDF/A-2 dispone di alcune funzionalità che erano precedentemente vietate, come l'utilizzo dell'algoritmo di compressione JPEG2000, l'utilizzo degli effetti di trasparenza, l'utilizzo dei *layer*, etc. Altre funzionalità rimangono, invece, vietate, come l'utilizzo di codice Javascript o l'utilizzo della crittografia applicata all'intero file.

Ai fini pratici è bene avere chiaro il fatto che il PDF/A-2 non sostituisce il PDF/A-1 né lo rende obsoleto; infatti, un file conforme alle specifiche del formato PDF/A-1 è comunque conforme anche alle specifiche del PDF/A-2 (ovvero è “PDF/A-2 compliant”). Di conseguenza, in tutti quei casi in cui per la conservazione dei documenti è già stato adottato il PDF/A-1 e si ritiene che le sue caratteristiche ben si adattino alle caratteristiche dei documenti da conservare, non c'è alcun bisogno di effettuare operazioni di conversione dei file PDF/A-1 già esistenti in file PDF/A-2 perché i file non verrebbero in alcun modo modificati e, quindi, da questa operazione non si trarrebbe alcun vantaggio.

Invece, in tutti quei casi in cui il PDF/A-1 non è una scelta ancora definitiva, è opportuno valutare se le nuove funzionalità del PDF/A-2 si adattino meglio di quelle del PDF/A-1 alla tipologia di documenti che si intende conservare. In questi casi, anziché “forzare” tali documenti ad essere conformi alle specifiche del PDF/A-1, con gli inevitabili adattamenti che tale operazione comporta, è opportuno prendere seriamente in considerazione la possibilità di definire una strategia di conservazione digitale che si basi sull'utilizzo del PDF/A-2, anche in considerazione del fatto che cominciano già ad essere numerosi gli strumenti in grado di supportare il nuovo standard²⁵.

²⁵ Sono già diverse le aziende che offrono prodotti software in grado di gestire il nuovo formato: ad esempio, Adobe, Callas software, Luratech, Adlib, etc. già hanno inserito il supporto al PDF/A-2 nelle nuove versioni dei loro applicativi.

Valeria CALABRESE

Gli archivi di impresa nel Biellese: dal censimento delle fonti al Portale degli archivi del tessile e della moda*

Abstract: *The article describes the project developed in the area of Biella for the identification and exploitation of the business archives in the textile sector. The web portal and the scenarios for their advanced development are examined in connection with the national project, implemented by the National Archives and the Associazione nazionale archivistica italiana, denominated “Archivi della moda del '900”.*

Introduzione

Il progetto di “Centro rete biellese degli archivi tessili e della moda” pone le sue basi nel censimento degli archivi aziendali in corso di progressiva realizzazione da parte della Soprintendenza archivistica per il Piemonte e la Valle d’Aosta e della Regione Piemonte, e nasce per presentare il patrimonio documentario delle aziende tessili biellesi al termine di un approfondito lavoro di ricognizione, ricerca e comunicazione. L’articolo si propone di presentare una delle principali attività progettate dal Centro rete, costituitosi ufficialmente nel gennaio 2010 presso la Provincia di Biella: il “Portale degli archivi del tessile”, strumento ideato per raccogliere, elaborare e rendere disponibile la cospicua mole di dati raccolta dal lavoro di censimento (e successivi interventi di riordino) di un primo lotto di fondi archivistici di aziende tessili svoltosi nel corso degli ultimi anni.

Il primo paragrafo riassume le vicende dell’industria tessile in provincia di Biella, dalla nascita delle corporazioni di lanaioli e tessitori del tardo medioevo al grande sviluppo di alcuni marchi della seconda metà del Novecento. A partire dal secondo paragrafo ha inizio il discorso sugli archivi d’impresa e in particolare sulla loro valorizzazione. Alla varietà di tipologie documentarie presenti al loro interno (testuali, iconografiche, audiovisive) va ad aggiungersi l’archivio del prodotto, da intendersi come funzione primaria dell’impresa, che ne rispecchia i percorsi imprenditoriali, le continuità e le cesure. La valorizzazione degli archivi d’im-

* L’articolo è tratto dall’omonima tesi discussa a Urbino nel mese di febbraio 2011, nell’ambito del corso di laurea specialistica in scienze archivistiche, librerie e dell’informazione documentaria.