



O PDF/A, suas versões e a posição do Brasil

No dia 28 de setembro de 2005 a ISO, International Organization for Standardization, aprovou um novo padrão que trata de arquivamento eletrônico de documentos, através da norma ISO 19005-1 *Document Management – Electronic Document File Format for Long-Term Preservation – Part 1: Use of PDF 1.4 (PDF/A-1)*. Esse padrão é o fruto de três anos de reuniões, discussões e revisões por organizações e empresas de diversas partes do mundo. Para se ter uma idéia do interesse gerado por essa iniciativa podemos retroceder a 2002 e observar os primeiros passos dados nesse sentido.

A primeira reunião para elaboração da norma, em 2002, já envolvia agentes de peso e com capacidade para impulsionar a iniciativa de maneira eficaz. Os envolvidos eram então a AIIM (Association for Information and Image Management), a NPES (National Printing Equipment Suppliers), o escritório administrativo das Cortes norte-americanas, Adobe Systems, Biblioteca do Congresso dos Estados Unidos, Surety Inc., Appligent, Merck, EMC, PDF Sages e NARA (National Archives & Records Administration). Em seguida, embarcaram no projeto a Xerox, Honeywell, EDS e Glaxo Smith Kline, entre outras empresas. Tais participantes redigiram o primeiro esboço da norma e submeteram seu projeto à ISO para ser registrado como um padrão internacional. Na ISO, o Comitê Técnico TC171 (Aplicações para Gerência de Documentos) ficou responsável pela norma com o auxílio do TC130 (Comitê de Tecnologia Gráfica), pois a norma é baseada no PDF/X, norma do TC130. O TC171 envolve 13 países participantes e 21 países observadores. Após revisões e numerosas emendas o padrão foi aprovado pela comissão da ISO.

Mas, o que significa essa norma? Qual será a extensão de sua aplicação e as implicações no mercado gráfico, nos arquivos das administrações públicas, autarquias, universidades e empresas? É o que iremos elucidar a seguir.

A EVOLUÇÃO

Na maioria dos países há normas quanto a arquivamento de documentos e, tradicionalmente, se trata de microfilma-

gem e microfichas. Porém, essa tecnologia está ultrapassada e já há, no atual estágio de desenvolvimento da informática, tecnologia boa e barata para substituí-la com enormes vantagens. Devido às dificuldades de se lidar com as microfichas, as empresas adotaram a tecnologia TIFF (formato de arquivo *raster*). Documentos são digitalizados e salvos no formato TIFF e postos para circular pelo mundo. Apesar de ser um formato estável, conservando as informações intactas, o TIFF, além de ocupar muito espaço de memória, elevando o tempo de transmissão pela Internet, não é adequado para pesquisas no interior das informações.

Com a evolução do PDF — que já está na versão 1.6, com o Acrobat 7 — o formato chamou a atenção do mercado como solução para as dificuldades que se tinha com o formato TIFF, especialmente ao adquirir uma face estável e neutra ao se tornar padrão internacional ISO 15930 (família PDF/X).

O PDF é um formato que pode conter texto, imagens vetor e imagens *raster*, além de *metadata* (metadados). Um documento deve ser digitalizado com a opção OCR (reconhecimento óptico de caracteres), pois assim, além da imagem do documento, o arquivo PDF conterá texto que servirá para buscas e índices. O PDF resultante ocupará apenas uma fração do TIFF correspondente, sem perda de qualidade, facilitando o envio por e-mail ou FTP. Os *metadados* podem ser incluídos para simplificar a posterior localização e organização das informações: título, autor, data de criação, datas de modificação, palavras-chave, etc. Sendo o PDF uma maneira de arquivamento independente das características de resolução e qualidade dos equipamentos de saída, o formato contém informação que se beneficiará dos avanços conseguidos em tecnologias de monitor e impressoras.

Para incorporar as vantagens do PDF, a flexibilidade do padrão PDF/X e as necessidades específicas do setor foi necessário criar, então, o padrão PDF/A.

O PDF/A

O padrão PDF/A é definido na norma ISO 19005-1 como “um formato de arquivo baseado em PDF, conhecido como

PDF/A, que prevê mecanismo para representação eletrônica de documentos de modo a preservar a sua aparência visual ao longo do tempo, independentemente das ferramentas ou sistemas usados para criar, armazenar ou render os arquivos”. Na norma não há orientação de como um sistema de arquivamento deva ser, nem estratégias de arquivamento a longo prazo, somente de como é o formato do arquivo e quais características de um PDF são admitidas para o PDF/A. O ponto central da filosofia do PDF/A-1 é que ele deve ser *autocontido*, isto é, todos os elementos necessários para se render o arquivo — mostrar na tela ou impressora — devem estar dentro do mesmo. Portanto, não há referências externas de fontes, imagens ou qualquer outro elemento.

PDF/A OU PDF?

Antes de mais nada é importante assinalar que o padrão PDF/A não inclui NADA que já não esteja presente em um PDF. Pelo contrário, a norma exclui elementos que possam ser um risco para os objetivos do padrão. Portanto, qualquer PDF/A é um PDF e NEM TODO PDF é um PDF/A. Alguns elementos do PDF são obrigados a existir, outros proibidos de existir e outros ainda são restritos caso a caso.

Por que o PDF/A não pode ser simplesmente um PDF? Quais são os elementos do PDF que se tornariam problema para os objetivos do PDF/A — repetibilidade, consistência, fidelidade, etc.? Os principais elementos são aqueles mesmos que foram afastados pelo pessoal do TC130 quando realizaram as normas PDF/X, que hoje reúnem uma família de normas: PDF/X-1a, PDF/X-2, PDF/X-3. Entre as restrições, podemos apontar as fontes não embutidas, imagens não embutidas, *links* externos, sons, filmes e ações javascript.

O PDF/A tem como base o PDF versão 1.4, aquela lançada com o Acrobat 5. A norma diz que “essa deve aderir a todos os requisitos do *PDF Reference* (da Adobe) da maneira como modificada por esta norma”. Dessa forma, o padrão identifica somente as diferenças com respeito à especificação da Adobe, que está publicada e disponível no site www.adobe.com. Para ser entendida a totalidade da norma é necessário entender por inteiro a especificação PDF.

PDF/A, A-1A, A-1B, A-2

A norma PDF/A é classificada como PDF/A-1 (Parte 1) e PDF/A-2 (Parte 2). O PDF/A-1 é posteriormente dividido nos níveis de conformidade PDF/A-1a e PDF/A-1b. Podemos detalhar essas normas e partes assim:

- ◆ PDF/A-1a (referido como nível A de conformidade) denota “conformidade total” com o padrão atualmente aprovado do PDF/A (ISO 19005-1): Parte 1.
- ◆ Há também um nível de “conformidade mínima” para o PDF/A: é o PDF/A-1b (referido como nível B de conformida-

de). Os requisitos do padrão PDF/A-1b visam a garantir sua reprodutibilidade idêntica e confiável ao longo do tempo.

O PDF/A-1a e o PDF/A-1b se diferenciam primariamente em relação à extração de texto. No PDF/A-1a é mantida e preservada a estrutura lógica do documento e o texto na ordem de leitura natural. O texto pode ser extraído e lido por computadores assim como equipamentos de mão como Palm e lidos em conformidade com a lei americana “*Section 508 of the US Rehabilitation Act*”. Já o PDF/A-1b visa a assegurar que o arquivo assim codificado possa ser rendido corretamente, em monitor ou impressora, mas não assegura que o texto que dele se extraia seja legível ou compreensível, portanto não atendendo ao quesitos da lei americana acima citada. Documentos com imagens digitalizadas ou imagens advindas de fax, SEM textos anexados, não sofrem diferença nenhuma se forem codificados na versão PDF/A-1a ou 1b.

Uma nova versão da norma PDF/A, parte 2 (PDF/A-2), está sendo concebida pela comissão do TC171 da ISO. Essa parte da norma visa a atender algumas das características adicionadas pela especificação 1.6 do PDF feita pela Adobe. Entre

Qualidade e respeito à vida

Acabamos de conquistar, de forma integrada,
a Certificação ISO 14001 e ISO 9001.
Para nós, isso representa mais
do que pioneirismo. Reforça nosso
compromisso com as questões ambientais
e com a satisfação de cada cliente.

Comemore mais essa conquista com a gente.



EXPOPRINT
LATIN AMERICA 2008

VISITE-NOS NA EXPOPRINT
STAND B35
LANÇAMENTO DO 1º CATÁLOGO
DE VERNIZES GRÁFICOS OVERLAKE

BRAINSTORM TCG



OVERLAKE
VERNIZES GRÁFICOS

O VERNIZ DA SUA GRÁFICA

Rua Colônia Leopoldina, 185 CEP 07220-040
Guarulhos, SP Tel. (11) 6488 0938
www.overlake.com.br E-mail: overlake@overlake.com.br



elas, poderemos ter *layers* (camadas), assinatura digital, *versioning* (versões) e transparência. Isso não afetará os arquivos criados segundo a norma PDF-1a, mas tornará certos arquivos PDF/A-2 não compatíveis com a norma PDF-1a.

O PDF/A E O BRASIL

No Brasil, a preocupação do governo com arquivamento de documentos não é recente e remonta à década de 70. É de grande importância, portanto, que as informações a respeito da norma PDF/A e as vantagens de sua utilização sejam divulgadas também entre os órgãos governamentais.

Para os que se interessarem em ter contato direto com a norma ISO 19005-1, ela pode ser encontrada no site da ISO (<http://www.iso.org>), sob forma de papel ou eletronicamente, no formato PDF. Está disponível somente em inglês. Espera-se que, em futuro breve, a norma seja traduzida e adotada pela ABNT, e possa fazer parte do ferramental das empresas, autarquias e governos. Isso certamente agilizará, no futuro, a compatibilidade entre os arquivos feitos no Brasil e os arquivos criados e mantidos pelos principais países ocidentais. Pensando-se nas enormes facilidades de integração, e conseqüente expansão de interesses diplomáticos, culturais e comerciais que isso promoveria, é que nos adiantamos em divulgar e promover o padrão para que a sociedade brasileira se integre no movimento global.

É claro que a leitura da norma exige um repertório anterior de conhecimentos profundos em PDF (versão 1.4), XML, RFCs, fontes, etc. Para que isso não pareça uma viagem frustrante diante da adoção do padrão, a ABTG, através de seu corpo de consultores, pode dar o apoio e suporte necessários para que a adoção do padrão PDF/A pelos interessados não se converta numa tarefa difícil.

O PDF/A E AS FERRAMENTAS

Ainda não estão disponíveis ferramentas de software para se implementar o PDF/A. O mercado prevê para breve o lançamento de programas, *plug ins* do Acrobat, e servidores capazes de configurar as partes de um sistema completo de fluxo de documentos a fim de alimentar um banco de dados PDF/A. O Acrobat 7 da Adobe já é capaz de salvar um documento como PDF/A, porém como foi lançado antes da aprovação final da norma, o resultado não é 100% garantido. A Microsoft anunciou que nas versões futuras de seus principais programas será possível salvar os documentos como PDF/A.

É claro que as grandes empresas e organizações terão que rever seus fluxos de documentos e procedimentos, sendo essa a base para a implementação vitoriosa de qualquer sistema de arquivamento. No caso da implantação de sistema de arquivamento de documentos com arquivos digitais

PDF/A o raciocínio é o mesmo: o fato de se arquivar documentos nesse formato não assegura a durabilidade ou se as informações serão vistas como se previu. A indústria tem consciência também de que o arquivamento de documentos em PDF/A não será a melhor opção em todos os casos. Mas, uma vez que se tenha decidido usar PDF/A, seguir a norma e procedimentos corretos garantirá a viabilidade do arquivamento digital de longo prazo. Entre os procedimentos corretos podemos citar fontes confiáveis, gerenciamento de qualidade e uma cuidadosa estratégia de migração dos arquivos TIFF para PDF/A, quando isso for necessário.

O PDF/A E O MERCADO GRÁFICO

São inúmeras as possibilidades que a formação de um gigantesco banco de dados proporciona para a indústria gráfica. Em primeiro lugar, tal acervo de informações é a base para que serviços de extração e impressão de dados customizados sejam viabilizados. Em segundo lugar, as universidades, editoras e escolas terão uma quantidade de informação jamais vista antes e necessitarão de “ajudantes” ou “agentes facilitadores” para poder circular entre essas galáxias de informação sem se perderem. Imaginar que arquivamento sob forma digital vá diminuir ou eliminar as formas impressas é tão ingênuo quanto aqueles que acharam que o advento de computadores, redes e Internet nas casas e nos escritórios fariam as impressões em papel diminuir. Elas só aumentam. Os produtos gráficos, portanto advindos de tal processo, podem ser imaginados como extratos, índices, catálogos e quaisquer outras maneiras de organizar essas informações sobre papel. No âmbito da ABNT, o órgão que naturalmente se preocuparia com uma norma dessa natureza seria o CB-14, que se refere à informação e documentação. Seu âmbito de atuação é “Normalização no campo de finanças, bancos, seguros, comércio, administração e documentação, compreendendo práticas adotadas em bibliotecas, centros de documentação e de informações, no que concerne à terminologia, requisitos, serviços e generalidades”.

Portanto é importante que a ABTG, enquanto licenciada para operar o ONS27, se antecipe aos outros organismos de normalização setoriais interessados a fim de capitanear o processo de nacionalização e divulgação da norma, garantindo assim a harmonia com as demais normas do setor gráfico e interesses advindos de sua implementação para o mercado gráfico.

BRUNO MORTARA é consultor da ABTG, sócio do estúdio de finalização Prata da Casa e coordenador da Comissão de Estudos de Pré-Impressão no ONS27, representando o Brasil no WG2 (Intercâmbio de Dados Digitais) do TC 130, grupo de trabalho da ISO.