

Petit guide pratique
d'utilisation du
PDF/A

Depuis deux ans, un format de fichier fait beaucoup parler de lui dans le domaine de l'archivage à long terme des documents numériques : **PDF/A**.

FedISA et SEAL Systems se sont associés pour rédiger ce document qui se veut un petit guide pratique à l'usage de ceux qui sont intéressés par la mise en oeuvre de ce format dans des projets d'archivage.

Ce document est destiné à répondre
à trois questions fondamentales
au sujet de PDF/A :

Pour qui ? Pourquoi ? Comment ?



FedISA (Fédération ILM Stockage et Archivage) est une association professionnelle qui s'est fixée pour principales missions :

- D'informer et assister les utilisateurs sur les nouvelles technologies, les nouveaux processus et les nouvelles réglementations en effectuant une véritable veille tant technologique que juridique, normative ou encore organisationnelle.
- D'aider les responsables de projet à justifier leurs dossiers : en les sensibilisant aux nouvelles technologies et aux obligations correspondantes, normatives, organisationnelles, juridiques et réglementaires, en leur permettant d'estimer les risques encourus (légaux et financiers) et les avantages compétitifs comme une meilleure réactivité.
- De fluidifier et participer au développement du marché de l'archivage électronique.
- De favoriser la mise en place de certifications et de référencements dans les domaines du tiers archivage, des outils matériels et logiciels, des systèmes internes d'archivage des entreprises.



SEAL Systems a développé un ensemble de solutions dédiées au PDF/A : **PDFLonglife**

Celles-ci permettent de vérifier et de générer des fichiers conformes à la norme PDF/A à partir de tous types de fichiers (MSOffice, graphiques, CAO). Ces solutions sont multi plates-formes (Windows, UNIX, Linux...) et facilement configurables. Les outils de supervisions apportent une transparence dans les traitements, un suivi complet (journalisation au format XML) et facilitent l'administration de la solution. Le traitement de fichiers en masse et l'intégration dans des workflows existants ont été éprouvés dans de nombreux projets. De plus, une intégration SAP certifiée et validée est disponible pour les utilisateurs de cet ERP. Une évaluation de la version interactive est téléchargeable en ligne : www.pdfa.fr ou www.pdfforever.fr. SEAL Systems est membre du PDF/A Competence Center.

Depuis la réforme du Code Civil de mars 2000, qui a introduit l'écrit numérique dans le droit français, la dématérialisation des procédures et des documents est devenue une réalité en France. Il suffit pour s'en convaincre de lire la liste des lois, des décrets, des instructions comptables du ministère des finances, des règlements, publiés depuis cette date.

Presque tous les domaines ont fait l'objet de travaux du législateur ou de l'administration : les ventes en lignes, les actes notariés, la gestion des appels d'offres des administrations, les services bancaires, les envois recommandés, les factures, le paiement de la TVA, les déclarations sociales...

Les avantages de la dématérialisation sont nombreux : gains de temps, réduction de frais de postes, possibilités de traitements automatisés, accessibilité à un document

parmi des millions en quelques secondes, pour ne citer que les plus évidents.

Mais la dématérialisation n'a pas que des avantages. Elle pose des problèmes techniques complexes : gestion de l'intégrité des documents, supports de stockage, moyens de sécurisation des documents sur le long terme...

Parmi tous ces problèmes techniques, le choix d'un format d'archivage est l'un des plus importants mais il n'est pas l'un des plus simples. Par exemple, il existe plus de 200 formats d'images fixes. Lequel choisir ? Il s'invente presque tous les ans un nouveau moyen d'encoder les séquences sonores. Lequel prendre ? Les formats des plans en 3 dimensions évoluent tous les 18 mois... Que retenir pour l'archivage ?

PDF/A est l'un des moyens de répondre à ces questions.

PDF EN QUELQUES MOTS

Avant de parler de PDF/A, il convient de préciser ce qu'est le format PDF.

Le **PDF** (Portable Document Format) a été initialement créé par Adobe Systems, pour gérer ses échanges de documents en interne.

Le PDF est un langage de représentation de documents, riche en fonctionnalités qui n'ont cessé de s'étoffer au travers de ses versions successives.

Les documents PDFs peuvent être simples ou complexes, structurés ou non structurés. Ils peuvent contenir du texte, des images, des graphiques et même des



contenus multimédia (documents sonores ou images animées). Les annotations, les métadonnées, les liens hypertextes, les signets ainsi que la signature électronique sont également supportés par ce format de fichier.

La documentation de la version 1.3 a été publiée en 1999, lors du lancement de la version 4 d'Acrobat Reader. Celle de PDF 1.4 a été publiée en novembre 2001, en association avec la version 5 d'Acrobat Reader. C'est cette dernière qui a servi de référence pour les normes PDF/A et PDF/X¹. Depuis, trois autres versions sont venues s'ajouter.

¹ PDF/X : Cette norme (ISO 15930) est destinée à assurer l'échange de documents dans le domaine de la pré-impression. Il permet notamment de piloter de façon homogène les RIP (Raster Image Processors), les logiciels qui permettent le tramage pour l'impression.

PDF a connu, jusqu'à ce jour, 5 versions depuis sa première publication :

■ **Version 1.3 - 1999** : PDF 1.3 intègre complètement le niveau III de PostScript. Il supporte la séparation des espaces de couleur et les profils ICC (profils techniques des numériseurs et des imprimantes);

■ **Version 1.4 - 2001** : Cette version introduit la transparence et les TAGs (gestion de la structure des fichiers) ; Les fichiers en niveaux de gris peuvent être compressés suivant la méthode JBIG 2 ; La sécurité est améliorée;

■ **Version 1.5 - 2003** : Avec cette version, JPEG 2000 peut être utilisé comme moyen

de compression des images fixes ; Par ailleurs, il devient possible de gérer des fichiers qui contiennent des couches²;

■ **Version 1.6 - 2005** : Cette version supporte les "open-type fonts" ; La longueur des pages peut atteindre 381 km ; Il est possible de gérer les fichiers qui représentent des objets en 3 dimensions;

■ **Version 1.7 - 2006** : L'encodage des documents est réalisé en UNICODE (tous les alphabets et tous les idéogrammes de la planète peuvent être employés dans les documents) ; Plusieurs fichiers PDFs peuvent être intégrés dans un seul fichier³.

QU'EST-CE QUE LE PDF/A ?

PDF/A est une famille de normes ISO (ISO 19005) qui définissent des sous-ensembles du format PDF à des fins d'archivage à long terme de documents électroniques.

La première de ces normes PDF/A-1 est le résultat de plus de trois années de travaux d'un groupe de travail de la commission 171 de l'ISO⁴.

Cette norme définit les exigences pour la création de documents adaptés à l'archivage, en se basant sur le format de fichier PDF. Elle spécifie en détail quels sont les types de contenu autorisés et ceux qui ne le sont pas. Ces spécifications sont destinées à assurer la lisibilité dans le

temps des fichiers, indépendamment des applications et des systèmes d'exploitation dans lesquels ils ont été produits.

PDF/A-1 repose sur la version 1.4 de PDF. Pour assurer la portabilité dans le temps, certaines fonctionnalités contenues dans le PDF 1.4 sont exclues de la norme, comme le mode transparent, l'inclusion de séquences sonores, d'images animées, d'images fixes comprimées avec JPEG 2000, etc. En revanche le PDF/A, impose certaines fonctionnalités comme l'incorporation des polices de caractères nécessaires à l'affichage des textes, l'ajout des métadonnées au format XMP (cf. page 6), etc.

2 Les couches (layers) permettent de gérer des superpositions dans un même fichier. Un des exemples classiques d'usage des couches est le multilinguisme. Imaginons la notice d'utilisation d'une machine : il y a une couche qui contient les photographies et les schémas de cette machine et il existe une couche de texte d'explication pour chaque langue (français, anglais, espagnol, italien, japonais, etc.). A la visualisation ou à l'impression, il n'est utilisé que deux couches à la fois, celle des photographies et des schémas et celle de la langue souhaitée.

3 PDF 1.7 est devenu une norme le 1^{er} juillet 2008 (ISO 32000)

4 ISO : International Standard Organisation

Différents formats de fichiers sont utilisés pour la conservation ou l'archivage des documents : les fichiers natifs, JPEG, TIFF, PDF, XPS... L'analyse des exigences fondamentales d'un format de fichier pour l'archivage à long terme montre que le PDF/A est un de ceux qui répondent efficacement à celles-ci. De plus, PDF/A apporte des avantages supplémentaires comparés à d'autres formats de fichier :

■ **Recherche texte avancée** : PDF/A peut enregistrer les documents en mode "texte", ce qui permet une recherche avancée sur un mot, des expressions, des chiffres. Les formats de type image (JPEG, TIFF, etc.) nécessitent l'utilisation d'un fichier associé qui contient le texte après OCR (Optical Character Recognition ou Reconnaissance Optique de

Caractères) afin de pouvoir être interrogeables de la même manière.

■ **Metadonnées** : Des métadonnées, comme le titre, l'auteur, la date de création du document, sa date de modification, le sujet, les mots-clés... peuvent être enregistrées directement dans le fichier PDF/A. Les fichiers PDF/A peuvent être ainsi classés automatiquement grâce à ces métadonnées sans intervention manuelle.

■ **Spécifications disponibles** : La publication de la norme garantit la possibilité pour toute entreprise ou toute administration de pouvoir accéder librement aux spécifications du format et donc pouvoir créer librement des logiciels de lectures ou de production de fichiers conformes à PDF/A⁵.

PDF/A : POURQUOI ?

PDF/A peut s'utiliser dans de très nombreux cas :

■ Brochures, manuels, fiches

d'information : De nombreuses sociétés et de nombreuses administrations produisent, diffusent et mettent en ligne leurs informations sous format de fichier PDF. Le PDF/A est un moyen fiable d'assurer l'**interopérabilité et la pérennité de ces informations**.

■ **Contrats électroniques signés** : Un nombre croissant de documents commerciaux est échangé de manière électronique. Les documents PDF/A peuvent être signés et faciliter la gestion de la valeur probante de ceux-ci en garantissant leur interopérabilité et leur préservation dans le temps⁶.

■ **Plans, cartes et dessins techniques** : Les cartes, les plans architecturaux, les plans de génie civil, les dessins techniques en ingénierie sont de bons candidats pour le PDF/A. Surtout lorsque la législation impose une conservation de plusieurs dizaines d'années pour certains de ces documents.

■ **Respect de la couleur des images** : Le PDF/A permet l'affichage correct de la couleur, élément essentiel pour l'utilisation d'images. Par exemple, pour les agences photographiques, les musées, les centres de

documentation, qui disposent de fonds importants d'images en couleur, le PDF/A est un moyen pour permettre l'archivage à long terme de celles-ci.

■ **Le monde médical** : L'imagerie numérique est de plus en plus répandue dans le monde médical. PDF/A est un moyen efficace et sûr pour l'archivage des dossiers des patients, des analyses, **des résultats de scanners...**⁷

■ **E-mails** : De plus en plus d'e-mails ont un caractère contractuel, chacun sait la difficulté d'archivage ou de migration d'un système e-mail vers un autre. Il est commode et simple d'archiver les e-mails en PDF/A.

■ **Documents pour l'impression** : Les sociétés d'édition, les cabinets d'architectes ou encore les graphistes peuvent utiliser la **compatibilité du PDF/X** avec le PDF/A. En effet, un fichier PDF peut respecter à la fois la norme PDF/X et la norme PDF/A, assurant ainsi pérennité et respect des couleurs.

■ **Interopérabilité** : Et, il ne faut pas oublier qu'un fichier PDF/A est aussi un fichier PDF... Il est donc lisible avec tous les visualiseurs de PDF. La représentation fidèle du document, quel que soit l'environnement logiciel, est donc ainsi assurée.

5 Cette caractéristique, essentielle pour un format de fichier destiné à l'archivage, est partagée par d'autres formats comme JPEG et JPEG2000

6 En matière de signature électronique, le format ne fait pas tout. Il convient de s'assurer que les méthodes, les outils utilisés et les procédures sont conformes à la législation en vigueur.

7 Attention, le format DICOM, très utilisé dans le monde médical, n'est pas pris en compte par PDF/A-1. En effet, ce format contient des images au format JPEG 2000, des images animées de type MPG2, etc. qui ne sont pas aujourd'hui supportés par PDF/A. Les futures versions de PDF/A devraient permettre l'archivage de ce format.

Un certain nombre de points clés sont à considérer lors de la conception et de la mise en oeuvre d'un projet d'archivage utilisant le format PDF/A :

■ **Images** : Toutes les images contenues dans les fichiers PDF/A doivent être reproductibles. Ceci peut être garanti uniquement par l'inclusion de l'image dans le fichier PDF/A (les liens vers un lieu de stockage externe au fichier sont interdits). Les images alternatives ne sont pas permises (une basse résolution pour l'écran et une autre HR pour l'impression). Les méthodes de compression autorisées sont le ZIP et JPEG. **Attention**, les compressions LZW et JPEG2000 sont interdites dans PDF/A-1. Le choix de la résolution des images ne fait pas partie du périmètre de la norme PDF/A. En conséquence, le choix de celle-ci est libre.

■ **Mode transparent** : La gestion de ce mode est imprécise dans la version 1.4 et peut influencer la représentation finale du document lors de son affichage ou de son impression. En conséquence, ce mode est interdit dans PDF/A-1. Par exemple, ces modes "transparent" peuvent être créés automatiquement dans une présentation PowerPoint puis inclus dans le fichier PDF. Il est donc nécessaire de faire très attention aux formats qui utilisent le mode transparent et qui doivent être transformés en PDF/A. Des tests préalables à leur conversion doivent être effectués.

■ **Couleurs** : Les couleurs d'une illustration ou d'un graphique dans un document doivent, normalement, toujours avoir la même apparence, quel que soit l'écran ou l'imprimante utilisée pour la visualisation ou la restitution. Le PDF/A répond à cette attente. Le PDF/A assure la gestion des couleurs grâce à différents profils disponibles (RGB, SRGB, Adobe RGB, ISO coated...). Cependant, selon l'utilisation du document PDF/A, un profil spécifique peut être inadapté. Son choix doit donc être effectué avec soin, même si ce profil peut être modifié dans le futur, pour tenir compte, notamment, des évolutions technologiques des moyens de restitution⁸.

■ **Polices de caractères** : Malgré la mise à disposition d'un grand nombre de polices de caractères usuelles dans les systèmes d'exploitation du marché (Windows®, Linux®, etc.) du fait du besoin d'affichage des pages Web internationales, il n'y a aucune garantie que les polices de caractères utilisées lors de l'affichage soient celles utilisées lors de la

création d'un document donné. Les documents PDF/A sont indépendants des polices de caractères de la plate-forme de restitution puisqu'ils contiennent toutes les polices de caractères qui leur sont nécessaires. Pour le niveau de conformité PDF/A-1a, les polices de caractères doivent être définies par référence à la norme Unicode.

■ **Metadonnées** : Les métadonnées sont des données intégrées dans un fichier et qui apportent des informations qui peuvent être exploitées par des fonctions de recherches. Dans le cadre des solutions d'archivage, elles sont utilisées pour une meilleure organisation et une meilleure gestion des documents. Pour PDF/A, les métadonnées doivent être encodées suivant le standard XMP (eXtensible Metadata Platform) basé sur XML. Ce standard a été développé par Adobe. Une seule métadonnée est obligatoire: le «PDF/A identifier», qui doit être produit par l'outil qui a généré le fichier PDF/A. Les métadonnées sont un point crucial à prendre en compte lors de la conception d'applications puisqu'elles fournissent les informations nécessaires au classement et à la recherche des documents archivés dans le temps.

■ **Fichiers PDFs Accessibles** : PDF dispose d'un ensemble de fonctions permettant l'accessibilité du contenu aux personnes ayant des déficiences physiques de la vue notamment. Les fonctions de grossissement et d'augmentation du contraste facilitent, par exemple, la lecture des textes pour les personnes ayant une vision diminuée. Les fichiers PDFs dits "accessibles" ont de nombreux points en commun avec le niveau PDF/A-1a. Les deux nécessitent la définition du contenu via des TAGs. Ces différents TAGs donnent des informations complémentaires sur le contenu et précisent ses usages. Les fichiers "accessibles" facilitent aussi l'affichage sur les petits écrans (PDA, Smartphone...) et les conversions éventuelles futures (PDF, HTML, TXT, RTF). Il est donc possible de créer des fichiers PDF/A-1a à partir de fichiers PDFs accessibles. Mais attention, le passage au format PDF/A doit être la dernière transformation de la chaîne de production.

■ **Fichiers PDFs Interactifs** : Les éléments interactifs améliorent de manière significative le périmètre des fonctions et d'usage des documents au format PDF. Par exemple, un fichier PDF peut contenir des hyperliens, des commentaires, des formulaires de saisie.

⁸ Cette possibilité de changer le profil colorimétrique peut induire des difficultés pour des documents signés puisque ceux-ci ne doivent pas être modifiés. C'est pourquoi le choix de ce profil doit être réalisé avec extrêmement de soin pour ce type de document.

Le PDF/A impose que les fichiers conformes à cette norme ne doivent contenir que des éléments insensibles au temps et qui n'entravent pas la clarté et la restitution des informations. En conséquence, la majeure partie des fichiers PDFs interactifs ne pourra pas être convertie en PDF/A.

■ **Commentaires et annotations** : Les commentaires ne doivent être ni cachés, ni paramétrés comme non imprimables. De plus, il y a certains types de commentaires interdits : «text edit comments», par exemple, car utilisant le mode transparent qui est lui-même interdit. Cependant, si un utilisateur souhaite mettre en place un commentaire permanent, il est possible de le faire mais uniquement via une note.

■ **Hyperliens** : Techniquement les liens hypertextes sont également des commentaires dans PDF. Ils ne peuvent pas être conservés en l'état. Ils doivent être «aplatis» (la référence du lien est présente mais le lien n'est plus interactif). Ces liens sont désactivés lors de la conversion en PDF/A.

■ **Formulaires** : La norme PDF/A n'interdit pas les formulaires, cependant certains types de champ utilisant des actions peuvent être incompatibles avec l'archivage long terme. Des impossibilités de conversion peuvent être rencontrées lors de la génération du PDF/A dans les cas suivants :

o **Présence de champs avec des actions du type «Soumettre un formulaire», «Importer données de formulaire», «Réinitialiser un formulaire»**. Ces actions sont interdites du fait qu'elles sont en mesure de changer le document.

o **Présence de scripts JavaScript**. Ces scripts sont également interdits puisqu'ils peuvent, eux aussi, modifier le fichier PDF lors de son utilisation ou de son affichage.

■ **Dessins techniques – CAO/DAO** : Les dessins techniques CAO, les plans (circuits électroniques, plans d'architectes, cartes...) sont de bons candidats pour l'archivage long terme en PDF/A. PDF 1.4 permet la gestion de formats de pages larges jusqu'à d'importantes dimensions (5,08 m X 5,08). Du fait des capacités de ce format, on peut donc concevoir soit un processus intégré de génération de PDF/A à partir de solutions CAO (Conception Assistée par Ordinateur) ou DAO (Dessin Assisté par Ordinateur), soit un processus de numérisation de documents papiers. Bien

évidemment, il est possible de marier les deux approches et d'avoir dans le même ensemble d'archives des documents issus d'outils de CAO/DAO et des documents issus de numérisations de documents sous forme papier. De plus, des métadonnées peuvent être mises en place, soit par reconnaissance de texte (OCR), soit par génération automatique, de manière à pouvoir utiliser les fonctions de recherche offertes par PDF. Cependant, les dessins en trois dimensions (dessin 3D) ne sont pas supportés par PDF/A-1 puisque ces derniers ne sont compatibles avec le format PDF qu'à partir de la version 1.6. En conséquence, s'il existe des dessins 3D dans une collection de plans à convertir en PDF/A, il convient vraisemblablement de réaliser un traitement préalable (par exemple, produire autant de fichiers PDF/A qu'il y a de couches dans le plan d'origine)⁹.

■ **PDF/A et signature électronique** : Les contrats commerciaux, les échanges administratifs sont désormais souvent effectués en ligne. Comment garantir l'authenticité des documents échangés ? Comme depuis des milliers d'années, la signature (manuscrite, sceau, tampon...) reste la solution utilisée sauf qu'ici elle est électronique. Les trois fonctions de cette signature sont de s'assurer de l'identité du signataire, de garantir que le document n'a pas subi de modifications depuis sa signature et que le document existait avec une date et une heure donnée. L'utilisation de la signature électronique avec les fichiers PDF/A ne pose juste que la question relative au séquençement des étapes du processus d'archivage des documents. Il faut d'abord produire le fichier au format PDF/A puis le signer. En effet, un document ne doit plus être modifié après sa signature. **Attention** : Ne pas confondre signature électronique et cryptage. En effet le cryptage (protection du contenu d'un document) est interdit pour la norme PDF/A.

■ **Des limites** : Dans certains cas, pour des raisons de sécurité, les documents doivent être resignés, après une certaine période, afin de pérenniser la signature initiale. Dans un fichier PDF, la re-signature conduit au remplacement de la signature initiale par une nouvelle du fait que la signature est embarquée dans le fichier PDF. Dans ce cas, il est peut être nécessaire d'utiliser un fichier de signature externe afin de garantir l'intégrité du fichier PDF/A et de ne pas être contraint à sa modification.

9 PDF/E (norme ISO 24517 Document management -- Engineering document format using PDF -- Part 1: Use of PDF 1.6) permet d'archiver les plans en 3D. Cependant comme PDF/E repose sur PDF 1.6, il faut attendre PDF/A-2 pour pouvoir l'utiliser pour l'archivage.

Les conseils simples qui suivent permettent d'éviter les erreurs lors de la production de fichiers au format PDF/A :

■ **Stratégie d'archivage** : PDF/A n'est qu'un format de fichier. En conséquence, l'utilisation de celui-ci n'empêche pas de se poser un certain nombre de questions : Quels sont les documents que l'on souhaite archiver ? Quelles sont les utilisations futures de ceux-ci ? Quels sont les types de documents, les durées de conservation, les volumétries présentes et à venir ? Par exemple, pour les documents papiers, il faudra considérer le temps de numérisation, le conditionnement (reliure, trombones, agrafes, etc.)... Selon les besoins, une fonction OCR pourra être mise en place pour permettre des recherches ultérieures en mode plein texte ou pour faciliter l'indexation... Autre exemple, la génération de fichiers PDF/A à partir d'e-mails doit être étudiée avec attention. En effet, les traitements peuvent être complexes compte tenu du type des fichiers attachés et de la récursivité de ceux-ci (un e-mail contenant un autre e-mail qui contient lui-même des fichiers comprimés, etc.). PDF/A ne sera un bon format d'archivage que si la stratégie d'archivage a été parfaitement définie et que si ce format a été retenu en tenant compte de ses caractéristiques propres.

■ **Fichiers protégés** : Les fichiers protégés par mot de passe (en lecture, en impression ou en modification) poseront des difficultés lors de leur conversion en PDF/A, voire même rendront impossible celle-ci. En conséquence, la sécurisation des fichiers (cryptage et mot de passe) doit être gérée au niveau du système d'archivage et non pas au niveau du fichier PDF/A.

■ **Polices de caractères** : Les polices de caractères utilisées dans les documents à convertir doivent être disponibles dans l'environnement logiciel de génération des fichiers PDF/A. En effet si, lors de la conversion en PDF/A d'un document, une police n'est pas présente dans l'environnement de production, la génération du PDF/A est compromise. L'idéal est que la génération du PDF/A soit réalisée par l'auteur du document qui lui dispose forcément de toutes les polices nécessaires à la production de son fichier. Si cette approche n'est pas possible (par exemple : producteurs en réseau et conversion sur un serveur), il est impératif de mettre en place des procédures strictes pour s'assurer que tous les polices utilisées par tous les producteurs de documents sont bien présentes sur le serveur de conversion / génération des fichiers PDF/A.

■ **Taille des fichiers** : Une augmentation de la taille des fichiers est souvent constatée (Les fichiers en PDF/A peuvent être de 5 à 10 fois plus volumineux que les fichiers PDFs originaux) lors du passage d'un fichier au format PDF vers un fichier au format PDF/A. L'une des causes qui peut expliquer cette augmentation de taille est l'inclusion dans le fichier PDF/A de toutes les polices de caractères nécessaires à son bon affichage. Une solution pour réduire la taille des fichiers est la concaténation de plusieurs fichiers en un seul car cette concaténation permet de mutualiser les polices de caractères pour plusieurs documents. Une autre solution est de maintenir dans le système d'archivage toutes les polices de caractères et de n'incorporer celles utilisées dans le document que lorsque celui-ci quitte cet environnement (dans ce cas les fichiers archivés ne sont pas conformes PDF/A). **Attention** ces deux solutions ne sont pas possibles dans le cadre d'un archivage à valeur probante (pour les documents signés) puisque les documents subissent une modification lors de la restitution dans les deux cas. Il n'est donc plus possible de valider leur authenticité.

■ **Compression** : Les compressions de documents sont possibles mais il est fondamental de noter que certaines compressions, comme la compression LZW, sont incompatibles avec les spécifications du PDF/A-1. Il convient donc de vérifier si les compressions qui seront mises en oeuvre par le logiciel de production sont compatibles avec le format.

■ **Vérification avant archivage : Est-ce bien un fichier PDF/A ?** Il est important de s'assurer que tous les fichiers PDF/A qui sont archivés sont conformes à la norme. En effet, une erreur dans le format peut rendre inutilisable un document dans 5, 10 ou 20 ans. Il convient donc de mettre en place un processus de vérification et de gérer les rejets en cas de non conformité par rapport à la norme. Il est aussi possible de ne pas rejeter les fichiers en anomalie mais de les normaliser lorsque cela est possible (fichier non protégé par un mot de passe, pas de polices de caractères atypiques...) par un logiciel adapté. Il convient donc de définir avec précision le workflow à mettre en place en considérant toutes les sources des documents et tous les traitements réalisés sur ceux-ci (cf. signature électronique). Quel que soit le mode de production (sur les postes des utilisateurs, sur des serveurs locaux, sur un système centralisé, etc.), il est plus que souhaitable (impératif) que le circuit d'archivage comporte obligatoirement une phase de vérification de la conformité des fichiers à la norme (voir ci-après "les outils").

Le format PDF/A comporte 2 niveaux de conformité :

■ **PDF/A-1a** : Le niveau le plus strict, puisqu'il assure la conformité de la structure logique du document à la norme. Chaque signe (lettre ou idéogramme) doit avoir un équivalent UNICODE. La structure et le contenu du document sont définis grâce à des TAGs. Ce qui permet une extraction précise du texte et une interprétation facilitée pour les appareils mobiles (PDA, téléphones...).

■ **PDF/A-1b** : Niveau suffisant pour les besoins de représentation. La structure logique du document n'est pas un élément essentiel, ce qui est important est la garantie de la représentation fidèle du document lors de visualisation par affichage sur écran ou impression.

Il est évident que lorsque, cela est possible, il est **préférable** d'utiliser le **niveau 1a** car ce niveau est le plus riche fonctionnellement. Cependant, ce niveau est plus délicat à obtenir puisque la structure du document doit exister dès la production de celui-ci pour être intégrée dans le PDF/A.

Pour résumer :

La génération du PDF/A-1b est plutôt réservée aux documents produits suite à une numérisation ;

La génération du PDF/A-1a est généralement réservée aux fichiers produits via des applications qui les produisent (traitement de textes, tableurs, etc.).

LE FUTUR : PDF/A-2

La norme ISO 19005-2 (plus connue sous le nom PDF/A-2) est en cours de rédaction. Comme la norme ISO 19005-1, la future norme ISO 19005-2 ne concernera que les documents statiques. Une troisième partie qui n'en est qu'au stade préliminaire (ISO 19005-3 - PDF/A-3) prendra en compte les documents interactifs (les documents comportant des formulaires notamment).

La norme PDF/A-2 sera basée sur les versions ultérieures à la version 1.4 (1.5, 1.6 et 1.7) du format PDF. La compatibilité ascendante sera assurée : si un fichier PDF respecte la norme PDF/A-1, il sera obligatoirement compatible avec la norme PDF/A-2. Plusieurs fonctionnalités sont en cours de discussion pour la version 2 de la norme (19005-2 / PDF/A-2), en voici quelques unes :

■ **JPEG2000** : Introduction de la méthode de compression JPEG 2000. Cette méthode est souvent plus performante que JPEG qui est aujourd'hui autorisée dans PDF/A-1.

■ **Mode transparent** : Jusqu'à la version PDF 1.4, la gestion des modes transparents et de la surimpression n'était pas clairement définie dans les spécifications de PDF. Pour cette raison, ces techniques n'étaient pas autorisées dans la norme PDF/A-1. Maintenant que ces modes sont clairement définis dans le format PDF, ils seront autorisés dans la norme PDF/A-2.

■ **Calques PDF (PDF Layers)** : Plusieurs vues pourront être enregistrées pour un même document (mais une seule peut être visualisée et imprimée à la fois : celle qui est affichée). Cette évolution est importante, particulièrement pour les fichiers CAO, car il sera possible de stocker des calques pour permettre la superposition de plusieurs plans dans un seul document.

■ **PDF Packages/Collections** : C'est une nouvelle fonctionnalité de la version PDF 1.7 qui permet d'embarquer dans un fichier PDF d'autres fichiers PDFs sans modification de ces derniers. Si certains documents doivent être regroupés, il suffit de créer un document PDF dans lequel ils seront conservés et d'ajouter les informations relatives à ces fichiers (les métadonnées) dans le fichier PDF les contenant. Comme les fichiers embarqués ne sont pas modifiés, les éventuelles signatures électroniques restent valides, l'intégrité du document étant préservée. Un fichier PDF "collections" sera conforme à la norme PDF/A-2 uniquement si tous les fichiers embarqués sont conformes à PDF/A-2.

■ **XRefs compressés et Streams** : C'est une nouvelle fonctionnalité introduite avec la version PDF 1.6. Les XRefs compressés sont utilisés pour les fichiers de grande taille (contenant de nombreuses pages et éléments) et assurent une amélioration de la compression de ces fichiers.

■ **Améliorations des TAGs et de l'accessibilité** : PDF/A-2 possédera de nouveaux TAGs qui faciliteront les accès aux documents.

■ **Utilisation native des polices de caractères Open type** : Les polices de caractères Open Type seront disponibles en mode natif. Ces polices sont d'un usage plus pratique que les polices Type1 ou TrueType.

■ **Amélioration des Métadonnées** : Les métadonnées embarquées sont plus riches que dans la version précédente de PDF/A et optimisent les accès aux documents.

■ **Signatures électroniques améliorées** : Une meilleure gestion des signatures électroniques sera disponible simplifiant l'archivage à long terme de celles-ci.

■ **En résumé** : La future version de PDF/A permettra principalement :

- De gérer des documents volumineux en rendant possible le regroupement de plusieurs documents dans un seul.

- D'améliorer les compressions des images (JPEG 2000 et gestion des couches).
- D'intégrer des fichiers 3D.
- De convertir avec beaucoup moins de restrictions les fichiers PowerPoint®.
- De faciliter la gestion des signatures numériques.

La norme ISO 19005-2 devrait être publiée au cours du premier trimestre 2010.

Pour résumer : PDF/A-1 peut être utilisé dès aujourd'hui pour l'archivage des fichiers issus de numérisations, de traitements de textes, de tableurs, etc. Si par contre, il est nécessaire pour une application d'archiver des séquences sonores, des images animées ou des images au format JPEG 2000, il convient d'attendre PDF/A-2 qui sera adapté à ce type de document.

LES OUTILS

Comme toujours, l'usage de bons outils est le gage de la réussite. Il convient donc de sélectionner ceux-ci avec soin.

Quatre types d'outils peuvent être utilisés :

■ Soit des outils permettant le contrôle de la conformité à la norme et aux niveaux de la norme (et dans le futur aux différentes versions de la norme) ;

■ Soit des outils de conversion de fichiers PDFs en fichiers PDF/A ;

■ Soit des outils de conversion de fichiers Offices, images, CAO en PDF/A ;

■ Soit des outils permettant la génération directe de PDF/A.

L'utilisation d'un logiciel de contrôle de la conformité à la norme est fortement recommandé. Il est en effet préférable de détecter immédiatement un problème plutôt que dans 10 ou 20 ans.

Pour une liste des outils et des fournisseurs, voir le site du PDF/A Competence Center.

POUR EN SAVOIR PLUS : PDF/A COMPETENCE CENTER

L'association internationale ADDS (Association for the Digital Document Standards) créée en 2006 est à l'origine de l'initiative PDF/A Competence Center, dont l'objectif est de promouvoir les échanges d'informations et expériences relatives à l'archivage long terme et à la norme ISO 19005 (PDF/A).

L'association organise et participe à des manifestations internationales, publie des ouvrages et organise des échanges au sein de son forum accessible sur son site web www.pdfa.org

Une présentation des membres de cette association et une liste de solutions logicielles autour de PDF/A sont également disponibles sur ce site.

CONCLUSIONS

Comme nous vous l'avons montré, PDF/A est un format bien adapté à l'archivage à long terme. Il suffit de connaître les quelques clés qui permettent de savoir parfaitement l'utiliser. Nous espérons que le présent document vous aidera à mettre en place efficacement vos applications utilisant PDF/A. Bien évidemment, nous mettrons à jour ce document au fil des évolutions du format PDF/A.

FedISA et SEAL Systems tiennent à remercier le *PDF/A Competence Center* pour la mise à disposition de sa publication «PDF/A in a Nutshell» que nous avons consultée lors de la rédaction du présent document.

FEDISA

55, avenue Victor Hugo
75116 Paris
France
Tél/Fax : +33 (1) 44 17 91 45
info@fedisa.eu - www.fedisa.eu

SEAL Systems

30 rue de la varenne
94100 Saint maur des fosses
Tel : +33(1) 43 97 69 66
info@sealsystems.fr – www.sealsystems.fr
www.pdfa.fr

PDF/A Competence Center

c/o LuraTech Europe GmbH
Kantstr. 21
10623 Berlin, Germany
Tel: +49 30 39 40 50 - 0
info@pdfa.org – www.pdfa.org
Référence : « PDF/A in a Nutshell » publié
par le PDF/A Compétence Center

ISO

1, ch. de la Voie Creuse
Case postale 56
CH -1211 Genève 20 - Suisse
Tél. + 41 22 749 01 11
Fax + 41 22 733 34 30
central@iso.org - www.iso.org/iso/fr

